Zipf's Law in L1 Attrition

A study of lexical diversity

Juan Miguel Vicente Flores

Supervisors:

Marjolein van Egmond
Sergey Avrutin

Second reader:

Gerrit Bloothooft

Thesis submitted in partial fulfillment of the requirements for the
Research Master's program
Linguistics: The Study of the Language Faculty

Utrecht Institute of Linguistics OTS
Utrecht University
Netherlands
July 2016

Zipf's Law in L1 Attrition

A study of lexical diversity

Juan Miguel Vicente Flores

Abstract

The present study is an attempt to connect the psycholinguistic study of lexical attrition and the different tools used in lexical statistics to measure lexical diversity, with a focus on Zipf's law and its potential as a lexical diversity measure. The theoretical introduction outlines certain theoretical explanations for forgetting language and vocabulary, and compares different tools used for measuring lexical diversity in previous research. Later, Zipf's law is presented, with a focus on its origin and the potential meaningfulness of its exponent as a measure of lexical diversity. Using data collected by Keijzer (2007), the spontaneous speech of a group of attriters, a group of controls, and a group of language acquirers was studied. Three research questions were investigated: whether attriter speech conforms to Zipf's law, whether it does so with the same or a different slope as controls (and acquirers) and whether the exponent is a strong predictor of group membership when compared with other measures of lexical diversity.

Using the same approach as in van Egmond *et al.* (2015), a good fit of the law was found in all three groups, with an unexpected better fit in the attriter group than in the control group. There were also significant differences in the exponent of the law among groups. The results in the exponent accord with previous findings in the study of aphasic speech and language acquisition, suggesting that the exponent might have linguistic relevance. The results are discussed relative to the different theoretical accounts of the origin of the law. To answer whether the exponent is a good predictor of group membership, it was compared with other measures of lexical diversity (number of types, number of tokens, type-token ratio, sample-independent type-token ratio, *vocd*-D, and *MATTR*), including the sociolinguistic variables of gender, education level and region of upbringing. The results show that the exponent failed to predict group membership of education level or region of origin, whereas sample-size independent measures like *vocd*-D performed better. This suggests that the exponent is not the most effective tool to measure lexical diversity.

Contents

Chapter 1

Theoretical Introduction

When speakers of a certain language move to a foreign linguistic environment and become proficient in the language of the new environment, they become bilinguals, and this change usually goes hand-in-hand with a magnification of both active and passive usage of their second language (L2), usually at the expense of usage of their mother tongue (L1). These speakers usually undergo some changes in their L1 that are known in linguistic studies as language attrition. Lexical attrition, the manifestation of this process in the bilingual speaker's lexicon, is believed to be the first step of the attritional process, and is manifested in the reduction of the speaker's lexical diversity. In this thesis, I present the different theories that explain this reduction. Later on, I explore the different methodological approaches that have been used to assess lexical attrition and I combine this knowledge with the use of Zipfian distributions to examine whether these distributions can be used to analyze lexical diversity and assess lexical attrition, and try to understand the meaning and relevance of the law itself.

1.1 Language Attrition

There is an important debate in bilingualism studies regarding the existence of attrition effects on the languages spoken by bilinguals. Broadly speaking, we can refer to language attrition when referring to "the decline of any language (L1 or L2), skill or portion thereof in a healthy speaker" (Ecke, 2004: 322). Today, there are two different branches in the study of language attrition. The first is concerned with first language (L1) attrition, and usually focuses on the linguistic characteristics of individuals that live for extended periods of time in a foreign linguistic environment. The second branch focuses on L2 attrition, and is a subfield of foreign language teaching. This thesis will be focused on L1 attrition, that is, the effects that are manifested in the native language.

This process of first language decline manifested by healthy speakers is usually experienced by migrants after moving to a foreign language environment for a prolonged period of time, and it has not been thoroughly studied in linguistic research until the last three decades. Nevertheless, it is still subject to some controversy regarding the similarities and parallels that this linguistic phenomenon shares with other linguistic processes such as language acquisition and language impairment.

L1 attrition can manifest itself at almost every level of linguistic knowledge, although it is generally believed that the first area to be affected is the lexicon, and the latest might be the phonology (for an extensive analysis on this topic see Schmid & Köpke, 2009). Phenomena of L1 attrition may include interferences from L2 to L1 in all levels (phonetics, lexicon, morphosyntax, and pragmatics), simplification of L1, and insecurity from the speaker when speaking their L1, to name a few.

Most authors agree that some of the factors that generally condition the apparition of language attrition are the age of acquisition of L2 and the length of residence in the L2 environment, the exposure time to both languages, the level of L2 proficiency, or the sociolinguistic characteristics that are linked to each language, i.e., the attitudes towards the native and the second language, and the habits regarding the use of these (e.g., Paradis, 2007; Pavlenko, 2004, 2010; Schmid & Köpke, 2007; Schmid, 2011, 2012, 2013). Language attrition is triggered by two factors. First, the speaker is in a foreign language environment and eventually develops a high proficiency in the L2 until it becomes the dominant language. This increased proficiency might alter the L1 system across all linguistic levels by means of interference. Such change has been referred to in the literature as *externally induced* change (Seliger & Vago, 1991). On the other hand, the input that the speaker receives in the L1 is drastically reduced, which might lead to *internally induced* language change (id.). Given that the native language is highly stable in mature speakers, it is generally believed that L1 attrition is an extreme and relatively rare development that only occurs under the confluence of these internally and externally induced changes, and after a prolonged time span: attrition phenomena in adult speakers are not visible in the first years after migration but take decades to be perceptible (Schmid & Köpke, 2009). However, some studies have found attrition effects soon after immigration, after less than five years (Köpke & Schmid, 2004; Pavlenko & Jarvis, 2002), while other studies found L1 systems to remain quite stable after decades of living in the L2 environment (Ammerlaan, 1996; Gürel, 2002; Hutz, 2004; Schmid, 2002), so the correlation between time of living in the L2 environment and attrition seems to be subject to some variation. Variation in experimental results has also been found when linking the frequency of use of the L1 to the degree of attrition (Schmid, 2013) and when relating attrition to a negative attitude towards the L1 (Cherciov, 2013).

More drastically, Laufer & Baladzhaeva (2015) found what they claim is L1 attrition without L2 acquisition: they found that attrition can occur in individuals after a long time living in an L2 environment without ever acquiring the L2, but just by interacting with other migrants in an (attrited) L1. However, these patterns can be explained attending language contact phenomena rather than using a language attrition paradigm. This relation between attrition, language shift and language loss has been studied in the literature (see, e.g., Hulsen, 2000). The language attrition paradigm used here, by definition (i.e., attrition being a phenomenon characteristic to bilingual migrants), relies on the changes that bilingual speakers experience in their L1 after gaining a considerable L2 proficiency caused by living in an L2 environment. The need for such proficiency in the L2 is essential in this psycholinguistic paradigm because it focuses on the interaction and competition of two linguistic systems. Köpke (2007) describes this phenomenon as a "restructuring of linguistic competence that entails a modification of the linguistic norms" (p. 20, cf. Py, 1986). How these changes may result in between-group interaction is an issue of sociolinguistics or language change rather than psycholinguistic studies, since they bear little resemblance to a psycholinguistic process.

1.1.1 Lexical attrition

It is generally believed that attrition affects the different components of linguistic knowledge at different rates and to different extents, but generally in the same order (Paradis, 2007). It is generally accepted that the lexicon is the linguistic area that is most susceptible to language attrition, and the first area where signs of attrition can be detected. This affirmation is almost axiomatic, since it has been widely reinforced in the literature (e.g., Andersen, 1982; Hulsen, 2000; Köpke, 2002; Köpke & Nespolous, 2001; Köpke & Schmid, 2004; Montrul, 2008; Opitz, 2011; Weinreich, 1979; Weltens *et al.*, 1993; for an overview of literature on the matter, see Schmid & Köpke, 2009). Other areas of language like morphology and syntax are less prone to erosion and only occur later in time, usually after decades of living in the L2 environment (e.g., Håkansson, 1995; Keijzer, 2007; Schmid, 2002; Hutz, 2004). One possible explanation for this phenomenon is that lexical items are not integrated into the structural system of the language (when compared to other areas like syntax or morphology), which makes them more susceptible to being lost: the losing of individual lexical items would not result in the restructuring of the lexicon, as would happen, for example, with morphological or phonological items. Another possible explanation for this is that given by the regression hypothesis (Jakobson, 1941), which claims that the order of language decline is opposite to the order of language acquisition. This would result in lexical attrition being triggered by regression: words would be forgotten first because they are learned after other areas of linguistic knowledge like phonology or syntax. Explanation of the regression hypothesis will continue in Section 1.1.4.

Lexical access difficulties in L1 have been found in attriters when compared to controls in numerous studies (e.g., Hulsen, 2000; Keijzer, 2004, 2007; Montrul, 2008). But how can a decrease in lexical accessibility be interpreted as a sign of language attrition? In late bilingualism, it is believed that the onset of attrition is the onset of bilingualism (Cook, 2005). Some authors believe that speakers who show lexical attrition will show, besides difficulties in lexical access activated in speech production, a difficulty in lexical recall in speech perception (Olshtain, 1989; Zelealem, 2003: 108; Sasse, 1992). But there are a range of phenomena that are common to both attriters and bilinguals in general, like tip-of-the-tongue states, weaker results in verbal fluency and longer response times in picture naming tasks. This has been argued to be a result of cross-linguistic interference arising directly from competing lexical representations and weaker links between semantic and phonological representations, due to the fact that bilinguals use the words of each language less often compared to monolinguals (Michael & Gollan, 2003: 395-397). This is better understood in the Activation Threshold Hypothesis (Paradis, 2007), that will be explained in Section 1.1.4.

Some authors consider code-switching a phenomenon characteristic to lexical attrition. Nevertheless, Du Bois (2015) analyzed a group of US Americans living in Germany and found that code-switching was a socially motivated phenomenon that could not be correlated to other developmental signs of lexical attrition. Code-switching is present in most bilingual communities, or might be present in conversations between bilingual speakers who, given the case, could show perfect lexical performance in both languages. Lexical attrition, on the other hand, manifests itself as a "forgetting" of lexical items in the mother tongue.

Lexical attrition phenomena include tip-of-the-tongue experiences, reduced verbal fluency performance and reduced lexical repertoires. As a result, attriters will have a reduced number of representations with a smaller variety of lexical items available to them, and those items retained would consist of more common and highly frequent items (Andersen, 1982), so word frequency is an essential aspect to look at when dealing with lexical attrition. Why the number of accessible elements is reduced will be discussed in Section 1.1.3.

1.1.2 Forgetting language

The first issue that needs to be addressed is the nature of the attritional process. Does L1 attrition mean (partially) forgetting one language? And what, exactly, is forgetting? This is a question that has been addressed both by (psycho)linguists and psychologists (e.g., McElree *et al.*, 2000; Yeni-Komshian *et al.*, 2000; Ecke, 2004). According to Ecke (2004), psychologists consider that there are three basic components to memory: encoding, storage and retrieval. In healthy individuals, the encoding component is usually assumed to be intact, leaving out storage and retrieval as the components that might be involved in the process of forgetting.

Another assumption that is generally made is that only what has been acquired can be forgotten. However, there exists the possibility that linguistic structures are not completely acquired, and the unstable storage of these structures may have contributed to poor performance and a sensation of forgetting (id.). This happens, for instance, to migrants that leave the L1 environment at a very early age, prior to having fully acquired their L1. This phenomenon is what linguists have called incomplete acquisition (Isurin, 2000; Levine, 1996; Montrul, 2002, 2008; Polinsky, 2006; Saville-Troike *et al.*, 1995; Sorace, 2005), and might result in language loss. Intuitively, it can be said that the more the attriters learn before migrating to the L2 environment the less they will be affected by attrition. This has led some authors to hypothesize that there exists a critical threshold for linguistic knowledge that, if reached, will make them immune to interference or decay. This is known as the *critical threshold hypothesis* (Neisser, 1984; de Bot & Clyne, 1989), and it is widely reported in L2 attrition studies (e.g., Lee & Schallert, 1997; Tomiyama, 2000; Weltens *et al.*, 1993, 1989), but also in some cases of early L1 attrition (Pan & Gleason, 1986; Yukawa, 1997). Furthermore, this would predict why younger migrants are more susceptible to L1 attrition, since it is more likely that they did not reach the threshold before departing to the L2 environment.

Köpke (2007) analyzed the different brain mechanisms and cognitive processes involved in language that may have a repercussion in the nature of the development of language attrition. She claims that attrition depends on four main *brain mechanisms*. The first one, *plasticity*, is said to be "a function of the age of onset of bilingualism and/or attrition" (Köpke, 2007: 6). The idea of plasticity is based on the fact that synaptic connections are not fully mature in the first years of life, meaning that younger migrants can adapt to the new language environment quicker than older migrants. Plasticity also predicts stronger L1 attrition in younger migrants, and this, together with incomplete acquisition phenomena, help corroborate the hypothesis that age is one of the most predictive factors in attrition. The second brain mechanism involved in language attrition would be *activation thresholds*, which can be seen as a function of how frequently a language is used. This mechanism will be discussed later since it is the key element in Paradis' Activation Threshold Theory

(Paradis, 1993, 2004). The third mechanism, *inhibition*, is said to be a function of type of language use: there are some inhibitory neural cells that help control neural activity, which are responsible for inhibition mechanisms that, together with activation, help achieve language selection in bi- or multilingual individuals, and can be used to explain phenomena like interference, processing difficulties or difficulties in L1 production (Köpke, 2007: 4-5). Lastly, the brain mechanism of *subcortical involvement*, is seen as a "function of emotional implication" (id.: 7). This mechanism is responsible for cases of attrition that involve traumas (e.g., Schmid, 2002; Isurin, 2000; Nicoladis & Gabrois, 2002; Saville-Troike *et al.*, 1995), but it can also be predicted that emotion is a key factor in any case of attrition, although this has not been thoroughly studied. Regarding the cognitive processes involved in attrition, Köpke claims that long term memory should be considered first, since it is where linguistic knowledge is stored over time, and any information stored in the long term memory that is not regularly activated might be forgotten. But what exactly is forgetting?

Ecke (2004) claims that both psychological and linguistic studies should aim to answer whether forgetting implies the loss of information or it is just the result of an impaired access to information. He distinguishes seven different processes: decay, interference, regression and suppression, distortion, retrieval slowdown and failure, cue dependency and interaction, and dynamic systems. Many of these concepts are difficult to distinguish because they are likely to be interrelated, and when attrition is analyzed from a psychological framework, there is no one process that can be said to trigger the whole phenomenon, but rather different processes that interact with a multitude of internal and external factors. In the next section I address whether lexical attrition is the result of a loss of information or, rather, an impairment of access to that information.

1.1.3 Forgetting words

When speaking about lexical access, one could distinguish between two types of memory mechanisms: one would be recall, which is the ability to produce a word (and involves activating a semantic form to retrieve a phonological form); while the other would be word recognition, which is the passive recognition of a word and its meaning without having to retrieve the phonological form (but rather involves accessing the semantic form of a given phonological form). Word recognition requires a smaller amount of cognitive effort than recall. Some authors (e.g., de Bot *et al.*, 2004: MacLeod, 1976) make the assumption that words cannot be completely lost once learned, and even when speakers are not able to recognize a word, they have some residual knowledge (savings) that can be reactivated through relearning. In a savings paradigm, they usually compare how speakers recognize old words (that are assumed to leave some residual knowledge) versus new items (that are assumed to be completely new) after a learning session. The expected outcome is that old vocabulary reaches the recognition threshold better than newly-introduced items.

MacLeod (1976) investigated the representation of meaning and input language in bilingual memory. Using the savings paradigm, he focused on the kind of residual information (phonological form, semantic information or information of the language of the word) contained in memory traces of the L1 and L2 words that could not be remembered five weeks after the onset of the study. He found that there were significant savings for *same* versus changed meaning but not for same versus changed language. MacLeod claimed that the results suggested that translation equivalents do not function as synonyms but are mediated by an underlying supralinguistic concept. But they also suggest that, over time, meaning is retained more easily than form and language information.

A fair question to address when wondering how words could be lost is to wonder how these words are stored. Schmid & Köpke (2009) claim that not much effort has been made to understand lexical attrition within bilingual lexica models. Bilingual lexica models can be of two types: models that separate lexica for each language, and connectionist models, that claim that there is one bilingual (or multilingual) lexicon and the items inside are sorted by language by means of activation and inhibition mechanisms. One of the models that follow the first type is the Revised Hierarchical Model (Kroll, 1993; Kroll & Tokowicz, 2001; Kroll *et al.*, 2005; the last two also for an overview of previous models of lexicon in the literature): in this model there are two differentiated lexica, with links that are established between the lexica with varying strength (depending on the proficiency or language use). When words are not used, the links can weaken, and this might lead to forgetting lexical items. An example of a connectionist model would be the Bilingual Interactive Model of Lexical Access (Grosjean, 1997). According to this model, activation and inhibition mechanisms would regulate the crosslinguistic links between items in a shared lexicon that would include phonetic features, word forms, lemmas, concepts or a language subsystem associated to every item. These links would also be dependent on frequency of use. It seems that, in these models, forgetting is seen as the weakening of the connections between the elements of the lexicon, and not the disappearance of those elements. The weakening of these links is directly related to frequency, making small frequency of use the trigger of forgetting an L1 or L2 form, regardless of whether that form is part of a unified lexicon where all L1 and L2 words are together, or rather words are arranged in separate lexica, according to whether they conform the L1 or the L2. But what happens to the phonetic representation of the word after the link has been completely weakened?

The forgetting of phonological forms might be explained by two of the theories of forgetting explained by Ecke (2004): retrieval slowdown and failure, or decay. What both these processes share is that frequency of use is crucial for the maintenance and access of information in memory. *Decay* is the oldest approach, and it claims that information vanishes or declines gradually in memory through lack of use (Thorndike, 1913). The lack of use of the information would result in the dissipation of a "trace" that has been imprinted for a piece of information represented in the brain (Ecke, 2004: 325). Retrieval slowdown and failure, on the other hand, claims that forgetting does not have to represent the loss or erasing of information from memory, but forgetting might rather be a problem in achieving access to the desired information (originally in Ashcraft, 1989; Spear & Riccio, 1994). Information is stored according to frequency of access, such that more recently accessed information is accessed faster. Neural network and spreading activation models claim that retrieval slowdown and failure is a consequence of the decreasing activation levels in neurons and processing units and the weakening of connections between neurons and nodes (Burke, 2002).

Evidence for decay is found in elderly monolingual speakers who are not in a language contact environment: these speakers usually experience word finding problems, such as tip-of-the-tongue states, and they experience these more frequently than younger adults. These findings have been attributed to biological decline that leads to a weakening of the connections between semantic and phonological representations, known as the *transmission deficit hypothesis* (Burke, 2002). An alternative explanation would be that older people fail to retrieve words more frequently due to quantitative reasons: they just know more words than younger adults. This is known as the *incremental knowledge hypothesis* (Dahlgren, 1998), and it can easily be extended to bilingual speakers. Findings from L2 attrition also support the concept of decay (cf. Ecke, 2004: 332).

On the other hand, evidence for retrieval slowdown and failure can be found in the empirical results that point to forgetting not being a total erasing of information but a deterioration in the access to that information. The first evidence comes from productive skills usually being more affected than perceptive skills: word recognition rates are usually higher than word retrieval rates in attrited speech (e.g., Bahrick, 1984; Kaufman & Aronoff, 1989; Tomiyama, 2000; Waas, 1996). Tip-of-the-tongue states are also seen as evidence for retrieval slowdown and failure: speakers can feel like they know a word and feel close to retrieving it, without succeeding at first. The fact that they sometimes do succeed shows that these memory failures are not cases of loss of information, but failures of access (Ecke, 2004: 334). McElree *et al.* (2000) investigated the speed of conceptual retrieval where participants of three different groups (L1 dominant, L1-L2 balanced and L2 dominant) completed a timed semantic categorization task where they had to decide whether different pairs of words belonged to the same semantic category. They found a lower retrieval speed and a less accurate categorization for items from the nondominant language, even when this was the L1. These findings suggest that attrition involves a reduced retrieval speed of form-meaning mappings, i.e., an impairment in the receptive skills.

When it comes to lexical production and perception, retrieval slowdown and failure is one of the first and strongest indicators of language attrition, but sometimes the difficulty lies in distinguishing it from other phenomena. For instance, to distinguish retrieval slowdown and failure from decay (which also involves the deterioration of retrieval routes), it should be shown that "language structures are intact and that attrition is only a processing problem due to the temporary unavailability of retrieval cues or the interference of structures from a competing language" (Ecke, 2004: 334).

L1 attrition is often described as a selective process that affects different components of linguistic knowledge in a different order, at a different rate and to different extents (e.g., Keijzer, 2007; Paradis, 2007; Schmid & Jarvis, 2014; Sorace, 2005; Tsimpli, 2007). But the experimental approaches to this are usually based on performance studies that are not combined with lexical diversity in free speech. An exception to this are the studies by Keijzer (2007) or Schmid and Jarvis (2014). Evaluating attrition in such restricted tasks makes it difficult to understand if processes like decay or retrieval slowdown and failure are directly involved in on-time processing of lexical information. The use of natural language should be measured in situations where all linguistic levels have to be integrated in real time (Schmid & Jarvis, 2014). In these cases, the effects are expected to be more pronounced, and there are two possible explanations for a reduced performance: on the one hand, it can be that using two linguistic systems at the same time may lead to an increased cognitive load (incremental knowledge), which might result in processing problems.

On the other hand, it could be that lexical access problems arise from the increase in activation thresholds.

1.1.3.1 The Activation Threshold Hypothesis

One of the brain mechanisms explained by Köpke (2007) is the activation mechanism, used to explain monolingual and bilingual lexical access. Activation is seen as a neuronal basis of memory that is closely related to frequency of use. Paradis' Activation Threshold Hypothesis (Paradis, 1993, 2004, 2007) assumes that a certain amount of neural impulses is needed in order to reach the activation to access items stored in memory. The activation threshold needed to activate a certain memory item (i.e., the energy required to activate it) is determined by frequency and recency of use: frequently used items or items that were accessed more recently are easier to retrieve. This explains why attrition has a bigger effect on less frequent items, and particularly on people who make less use of their attrited L1 (Andersen, 1982; Paradis, 2007).

Another brain mechanism involved in determining the activation threshold is inhibition: when selecting an item from the lexicon, its competitors must be inhibited. This mechanism of inhibition will consequently raise the activation threshold of the competitors, so in the future, when they have to be activated again, they will require more cognitive effort (Paradis, 2004). Bilinguals routinely inhibit the language which is not chosen for activation, which may lead to word-finding difficulties in the inhibited language: in the case of L1 inhibition, this would lead to L1 lexical attrition (Köpke & Jarvis, 2014).

1.1.3.2 Conclusion

Summing up, we have seen that bi- or multilingual speakers usually conform to a language system that is different to that of monolingual speakers in that they have two (or more) language systems. How these systems interact is a matter of debate. On the lexical level, theories like decay claim that being dominant in an L2 may lead to the dissipation of lexical items in L1 because of lack of use. A different view would be assuming that these two language systems are active at the same time, and compete during lexical access, leading to bilingual processing requiring higher cognitive effort. Competing lexical items will be more easily accessed when they were used more frequently, earlier and when they were not inhibited. This results in the reduction of fluency and a slowdown in bilingual lexical access, but is also observable in interferences on the lexical, phonological, and grammatical levels in both languages.

1.1.4 Attrition, acquisition and breakdown

The knowledge and use of language cannot be seen as something constant or linear. This can be seen, for instance, in what is known as the *language reversion* phenomenon: after elderly migrants have reached a considerable level of attrition, the process can be reversed, in a way that their L1 fluency improves because of a switch back to the use of the mother tongue. This happens, for instance, in elderly speakers: sometimes, families where both parents spoke the L2 with their children, go back to speaking the L1 after their children emancipate (see, e.g., Schmid &

Keijzer, 2009; but also de Bot & Clyne, 1989, 1994; Fishman, 1991; Hermans, 2012; Hinton & Hale, 2001; Kapanga, 1998; Vakhtin, 1998). This is an example of how linguistic knowledge is continuously in flux: it is not necessarily linear, and it is dynamic in nature. Some of the possible outcomes of this dynamism are the acquisition of a (second) language, the (partial) loss or attrition of it, or the reversion phenomenon just mentioned. Learning a second language can affect negatively the native language system resulting in attrition, and in extreme cases or cases of early acquisition of the L1 it can lead to language loss.

Another example of the nonlinearity of the language process is what is referred to here as language breakdown. In some cases, language knowledge can be disrupted by brain injuries. A particularly well-studied case is that of aphasia, which is the language disorder resulting from stroke. One of the aspects of aphasic speech is that people often have difficulties in word retrieval. Some authors use the notion of activation levels in the lexicon to explain these difficulties, similarly to the different accounts that explain lexical attrition. For instance, Paradis tested the activation threshold hypothesis to account for differential inhibition in polyglot aphasia and also in healthy bilinguals, and found that the hypothesis held (Paradis, 1985, 1993, 2004). Following van Ewijk (2013), the explanations of lexical access difficulties in the literature can be grouped in three blocks. The first block of research claims that base-level activation of lexical representation has been reduced (e.g., Yee *et al.*, 2008), delayed or slowed (e.g., Swinney *et al.*, 2000), leading to slower-than-normal lexical activation. This explanation goes hand-in-hand with the theory of retrieval slowdown that explains lexical attrition. A second block of research claims that there are issues with aphasics' working memory that result in difficulties with lexical access (Martin *et al.*, 1999): aphasics would have "difficulty maintaining the activation of lexical representations in working memory, which leads to a failure in lexical access" (van Ewijk, 2013: 65). This theory would be parallel to activation failure. The last block of research claims that there is a problem with cognitive control, that would lead aphasics to have trouble selecting a target between competing alternatives, being unable to overcome lexical competition (Utman *et al.*, 2001). This is reminiscent of the incremental knowledge hypothesis.

Another interesting language pathology that involves impairment in lexical access is Alzheimer's Disease (AD). People who suffer from AD undergo linguistic alterations that can be observed in all linguistic levels. The linguistic abilities that are more resistant to decline are articulation and perception or reading (Patterson *et al.*, 1994). These two are also reduced in the most advanced stages of AD. On the other hand, vocabulary seems to be the linguistic level that is earliest affected by the disease (Pekkala *et al.*, 2013). People who suffer from AD have characteristic speech features: their speech is characterized as empty, because it includes a high proportion of words or expressions with little or no semantic content (Aronoff *et al.*, 2006; Almor *et al.*, 1999; Kempler *et al.*, 1987; Hier *et al.*, 1985). This empty speech often contains words that appear with high frequency within the text, mostly functional words like prepositions, articles, conjunctions or auxiliaries. Low frequency words such as proper names and non-frequent names are extremely affected (Patterson *et al.*, 1994; Shuttleworth & Huber, 1988). This characterization of AD speech, in terms of frequencies, is parallel to the speech of people with (severe) lexical attrition, who Andersen (1982) claimed would have a reduced number of lexical items, and those items retained would consist of more common and highly frequent items.

These similarities between language disorders and attriter speech has led many authors to consider language attrition to be the 'healthy language disorder'. Hence, recent research has tried to give a transversal account for language development in a holistic way. There is, for instance, an increasing number of studies that try to relate language acquisition theories to L1 attrition (e.g., Gürel, 2004, 2011; Keijzer, 2004, 2007, 2010; Sorace, 2005; Tsimpli *et al.*, 2004). Attrition has also been related to language pathology. Sometimes, attrition is used as a general term that includes both language disorders and L1 attrition. For instance, Dressler (1991) compared L1 attrition (which he called language decay) to aphasic speech. Some of the findings were that "both types of attrition present[ed] dysfunctional language production" and "both types of attrition show[ed] signs of insecurity in performance" (Dressler, 1991: 109). By "both types of attrition" he meant L1 attrition and aphasia. In this paper, the use of *language attrition* is reserved for the attrition experienced by healthy individuals, and *language breakdown* is used to refer to unhealthy/pathological speech. Two of the hypotheses that have been postulated in this direction are the regression hypothesis and Ribot's law. They both postulate that learning and unlearning are correlated processes. I will explain these two and continue explaining a theory that accounts for more variables in the language development process and perfectly portrays its nonlinear nature: the dynamic systems theory.

1.1.4.1 Ribot's Law and the Regression Hypothesis

An important issue related to language forgetting is the question of whether there is a predictable order or sequence that can be predicted for language attrition. In the study of linguistic disorders, a main goal of research is to correlate the lost linguistic capacities to the areas of the brain that are injured, in order to corroborate or rule out the lastest findings in linguistic theory. Researchers have observed that some of the linguistic patterns of some aphasic speech was comparable to the linguistic patterns of children during the language acquisition process. One of the first psycholinguistic theories to tackle these similarities was Ribot's law. According to this theory, established for retrograde amnesia, there is a gradient, such that recent memories are more likely to be lost than more remote ones. The linguistic consequence of this would be that linguistic structures that are acquired later during language acquisition are more susceptible to being lost during language breakdown (Ribot, 1882; see also Avrutin *et al.*, 2001). Other authors have hypothesized that the complexity of the structures being learned interacts with the order of acquisition: the most complex or difficult-to-process structures that are acquired later are lost first, and that the less complex, easy-to-process structures that are acquired early, largely remain resistant to loss (Bailey, 1973; Pfaff, 1981; Slobin, 1975; cf. Ecke, 2004).

A more linguistically developed theory is the *regression hypothesis* introduced by Jakobson (1941), which predicts that the process in which a language is forgotten would be the reverse of the process in which the language was learned. Jakobson compared early stages of language acquisition with pathological language loss in aphasia. Particularly, he looked at the phonological system of Slavic languages, particularly plosives: the acquisition of phonological distinctions in point of articulation (labial, dental, velar) was acquired prior to voicedness distinction, i.e., the distinction between /p/ and /t/ was acquired before the distinction between /p/

and /b/. In aphasic speech, the order of dissolution was reversed: they lost the voicedness opposition before the distinction in point of articulation.

A number of studies have tried to compare the order of language acquisition with patterns in language breakdown. Keijzer (2007) claims that these attempts (i.e., Caramazza & Zurif, 1978; Grodzinsky, 1990; Avrutin *et al.*, 2001; Kolk, 2001; Bastiaanse & Bol, 2001) have failed to provide evidence proving that the order of language breakdown mirrors the order of acquisition due to three main problems. First, aphasia is the result of a brain injury which usually leads to *partial* language impairment, whereas acquisition is a global process that affects the whole language system. Second, aphasia is manifested as a sudden disruption of linguistic performance, whereas acquisition is a gradual process with no sudden improved performance. And third, aphasics usually show adaptation strategies that might be related to a restructuring of the neural substrate, strategies that are not comparable to language acquisition processes (Keijzer, 2007: 3-4).

1.1.4.2 Regression and attrition

The fact that attrition is a more gradual process than aphasia led researchers to wonder if attrition would be the process that could actually mirror language acquisition. Some studies focused on trying to find parallels between L2 acquisition and L2 attrition (e.g., Cohen, 1986; Hansen, 1999; for an overview see Ecke, 2004; and Keijzer, 2007), and results seem to point to structures that are learned last being the ones to be first forgotten. In comparing L1 attrition to L1 acquisition (e.g., Håkansson, 1995; Schmid, 2002; for an overview see Keijzer, 2007) the results were initially more problematic. Firstly, because there is no consensus in the literature about the patterns in L1 acquisition, the studies lacked the explanatory power for the regression patterns. And secondly, they focus on the patterns of early acquisition, and L1 attrition can only be compared to the fraction of the acquisition process of almost fully-mature acquirers (Keijzer, 2007).

Keijzer (2007) studied these mirror symmetries in a group of first-generation Dutch migrants in Anglophone Canada using a control group of Dutch non-migrants and also a matched group of Dutch adolescents. She found that there were symmetries between the loss of morphology and syntax when comparing the migrant speech to the adolescent group, providing a strong piece of evidence for the regression hypothesis. However, she claims that the evidence is not sufficient, and that research should not focus on establishing parallels in the order of language learning and attrition, but should also try to give it explanatory power by integrating it in a suitable theoretical framework (Keijzer, 2007; 2010).

Another hypothesis that can account for the order in which elements of a language are learned and forgotten is a generativist approach: the Interface Hypothesis (Sorace, 2005), originally formulated for L2 acquisition. According to this hypothesis, when speakers are learning a language, acquiring linguistic knowledge of a particular linguistic area is not problematic. Interfacing between any of those modules is not problematic either, but the problem would arise when the speaker needs to make grammatical operations that involve interfaces between an internal component of grammar (i.e., phonology, syntax or semantics) and an external one (i.e., pragmatics or discourse). The reason for this difficulty would be that purely syntactic features are more stable than features that are in the external interfaces. This would also apply to L1 attrition: the internal components of grammar would be more stable than pragmatic or discourse level information, so operations between internal and external interfaces (e.g., syntax-discourse) would be more vulnerable to residual or emerging optionality (Sorace, 2005; Gürel, 2011), and hence predicted to be lost first.

1.1.4.3 Dynamic Systems Theory

There are different theoretical frameworks that cover language acquisition, language attrition, and language breakdown. Keijzer (2007) focuses on the regression hypothesis, and discusses the explanatory power of previous frameworks, grouping them in three blocks: generativist explanations, usage-based explanations and Dynamic Systems Theory (DST). The problem with the first is that they rely on the existence of a Universal Grammar (UG) that is not yet fully characterized, and does not account for regression in a grammatical level since its UG constraints cannot be operationalized in language de-learning stages. Furthermore, it does not take into consideration external variables (e.g., language exposure, length of exposure, etc.) that have been proven to have an effect in language attrition, language acquisition or, by extension, phenomena resulting from the nonlinearity of language. Usage-based theories, on the other hand, see the language acquisition process as a result of cognitive processing of external input, which happens as a means of social interaction by way of intention reading and pattern-finding. Both acquisition and attrition can be explained with three mechanisms: analogy, entrenchment and preemption. Analogy, the ability to group perceptually and conceptually similar objects, provides acquirers with pattern-finding skills or pattern reading, and helps explain attrition for late acquisition: structures that are acquired later may not have reached the level of exemplars and cannot be generalized to novel contexts. Entrenchment and preemption explain competition of forms that might result in interference in attriters. During acquisition, verbs that are heard more often in a certain construction are less likely to be used in different constructions (entrenchment), but if the child hears adults use these verbs in structures that have a similar meaning, then s/he might readjust the use of such verb for such structure (preemption). This predicts that features that are entrenched will be more difficult to lose than preempted ones (Keijzer, 2007: 16-29).

However, these frameworks do not overcome the view of language development as a linear process with a clear beginning (like generativist approaches) or as a process with limited possible ending states (like usage-based approaches). Phenomena like disrupted acquisition, language attrition, language reversion, language loss or language impairment point to the need for a holistic theoretical framework that can account for these different directions in language development taking both external and internal factors into consideration. This has been attempted by applying complexity (or chaos) theory (DST) to cognition (e.g., Thelen & Smith, 1996) and language development (de Bot, 2007; de Bot *et al.*, 2007; Jessner, 2008; Herdina & Jessner, 2002). Dynamic systems are characterized by what is called *complete interconnectedness*, meaning that all variables included in a system are interrelated. When applied to language, languages emerge from the interaction of social patterns, patterns implicit in the input and pressures arising from general aspects of the cognitive system (MacWhinney, 1998). Dynamic systems are said to be highly dependent on the initial state, which means that small differences at the beginning, added to the nonlinearity of the systems, may result in huge differences in later stages of development (butterfly effect). De Bot *et al.* (2007) explains that, for instance, very subtle hearing problems in early childhood may result in long lasting difficulties in second language acquisition.

Dynamic systems non-linearity is characterized by the interaction of variables that affect each other over time – they change between stable and unstable states, they are *self-organizing*, and their order arises from self-organizing principles in nature. As they develop over time, the systems stop in *attractor states*, which are considered to be preferred, but still unpredictable. There are also states that are clearly not preferred, which are called *repeller states*. These concepts are used to explain the developmental patterns in language acquisition: usually children follow similar patterns in acquisition, although there are differences between acquirers. It also explains that deviant language repertoires (observable, for instance, in attriters or L2 acquirers) are systems in constant flux, and influenced by both internal and external information until they reach a steady (attractor) state.

In addition to de Bot *et al.* (2007), Larsen-Freeman (1997) has also applied DST to second-language acquisition. Herdina & Jessner (2002) applied DST to multilingualism and presented a Dynamic Model of Multilingualism (DMM). In their model, acquisition and attrition are seen as two possible outcomes of language development, which is presented as a continuum. Language attrition is seen as the result of not spending enough time or effort in the retention of a language, and it is directly related to competition with (an)other language(s). Both language acquisition and language attrition are seen as a competition between cognitive resources. The speaker's system will try to respond to this competition by self-organizing the resources until it arrives at an equilibrium. The source of competition in acquisition may be systems that are not fully mature (in interaction with the external input), whereas, in attrition, the source would be the interaction of the L1 and L2 competing for memory space. This is an interplay between system-internal resources (like intelligence, linguistic capacity, time or motivation) and system-external resources (like language and social input). It is usually the case that systems are in a steady state (like early stages of attrition), meaning that they will be reluctant to change, and when competition happens in those states changes will be very slow. But once a change has been triggered it will result in an acceleration of changes and the flux to the next attractor state (Herdina & Jessner, 2002; Keijzer, 2007).

In the dynamic view, when a speaker increases in L2 proficiency, their L1 proficiency is affected (Jia & Aaronson, 1999; McElree *et al.*, 2000; Segalowitz, 1991). Meara (2004) modeled the loss of vocabulary in language attrition. He designed lexical networks in which words were connected to two other words. Each element could be in an activated mode or in a deactivated mode. In his network, the activation of a particular word depends on the level of activations of the two words it is connected to, so the activation or deactivation of a word would lead to a cascade of activation/deactivation until the network settles in an attractor state. He varied the number of words that were active or not to study different attrition patterns, and found that when the number of deactivated words was slightly increased the effect in the pattern was remarkable: while sometimes it would lead to a clear attrition effect, other times the system remained stable. Although this artificial network cannot be compared to the human lexicon (in which lexical activation does not work under Boolean values), it represents how, in the same settings, there can be big individual differences. This would explain, for example, the differences found between attriters that live under similar circumstances: while some of them show signs of attrition shortly after moving to the L2 environment, others maintain the language in similar conditions to a monolingual speaker after more than 25 years (de Bot *et al.*, 2007).

To my knowledge, DST has also been applied to developmental language disorders, particularly to Specific Language Impairment (SLI, Verhoeven & van Balkom, 2003). Christman (2002) also tried to understand aphasic speech from a DST perspective. In her view, what differs between normal-developing and aphasic speech is that, while the first is normally chaotic (i.e., self organizing, locally unpredictable but globally stable), aphasic speech, as a result of an inflexible and asynchronous neurology, presents deficits that are explainable by an unstable (i.e., chaos-reduced) stage. She explains, for instance, that aphasics' reduction in speed of processing could be a result of the system approaching an attractor state. The reduction in the accuracy in lexical access, on the other hand, could be explained understanding the meaning of a word as the attractor space defined by all semantic microfeatures that are self-organized into one meaning. Aphasics would present a disruption in this self-organization, and the use of neologisms or circumlocutions would be a an attempt to overcome that disruption (and find another steady state).

1.2 Some Lexical Diversity Measures

Lexical diversity has been used as a diagnostic tool for clinical diagnosis of language disorders, for language proficiency tests, and as a tool to evaluate lexical attrition. When trying to assess lexical attrition, there are different measures that can be used to cover such phenomenon. The quality of these are usually judged in terms of how accurate they are in reflecting the complexity of the speaker's vocabulary knowledge. The main difficulties these measures need to overcome are sample-size effects and how well they predict language knowledge or behavior. In this sense, the first thing to tackle is the definition of lexical diversity, as opposed to other similar concepts like lexical sophistication or lexical richness. *Lexical diversity*, also referred to as lexical variability, lexical variation and lexical variety, refers to the "proportion of words in a language sample that are not repetitions of the words already encountered" (Jarvis, 2013: 88). Measures of diversity are inversely proportional to the word repetition rate. On the other hand, *lexical sophistication* measures the amount of used words that are not among the most frequent in language, i.e., the use of "sophisticated words" (e.g., using *summon* instead of *call*, or *request* instead of *ask*). Lastly, the use of *lexical richness* is more subject to ambiguity. Initially, it was used to refer to the number of words that a person knew, i.e., the "size" of the person's lexicon, but it is widely used as a textual measure, and used in the same sense as *lexical diversity*, so they will be used as synonyms here. Whilst some measures can give some insights into the lexical *sophistication* of a text, research has mainly focused in measuring lexical diversity.

In this section, an overview of the different attempts to measure this diversity is presented, following mainly Jarvis (2002, 2013), Malvern & Richards (2002, 2012), Schmid & Köpke (2009) and Schmid & Jarvis (2014). The explanatory power of each of the tools will be analyzed and commented on with lexical attrition in mind, and this will be accompanied by the most relevant findings in language attrition research and the study of other linguistic phenomena.

1.2.1 Verbal fluency tasks

Traditionally, lexical fluency has been measured with verbal fluency tests, and one of the favorite choices is the Verbal Fluency Task (VFTs). The underlying idea in these tasks is that participants focus all their attention on the lexical retrieval process (Schmid & Jarvis, 2014). The procedure consists in asking the speaker to name (and sometimes to write down) as many words as possible within different given lexical categories, usually in a fixed period of one minute. These tasks are easy to perform across languages since the linguistic material can be easily adapted. The responses are recorded and the answers that correspond to the cues given in the instructions are computed as valid. The score is calculated as the total *number of valid responses* (NVR) that were given in the pre-set time or counting the *number of different words* (NDW).

However, it has been pointed out that lexical productivity is also dependent on the choice of the semantic categories and on the design of the VFT. If the speakers are given certain categories with a big number of elements (e.g., animals, clothes), they score higher in these than in others with fewer elements (e.g., holidays). Researchers have tried to solve this by performing *formal* (versus *semantic*) verbal fluency tasks, by giving the speaker the word-initial letter as the only cue, but this criterion is still restricted to certain initial cues, since there are not as many elements that start with each letter of the alphabet. Nevertheless, formal verbal fluency tasks have been found to give rise to variation within normal populations (Roberts & Le Dorze, 1998) and to be sensitive to aging effects (Evrard, 2001: 182). Opitz (2011) claims that, in semantic tasks, both NVR and NDW are "influenced by the absolute size of the of a given stimulus" and "the familiarity of a person with a particular category", and usually the stimuli are chosen so that they are easy to retrieve, normally by means of common and ideally culture- and gender-neutral categories (Opitz, 2011: 113-14). This approach is, of course, problematic if we want to measure lexical attrition, since most theories claim that attriter speech is firstly characterized by the smaller use of non-frequent items, whereas the most common lexical items are usually unaffected even after showing other signs of attrition.

This type of task is widely used in neuropsychology, especially for patients with aphasia, Parkinson's and Alzheimer's diseases. They are also widely used as a diagnosis tool for Alzheimer's disease (e.g., Carnero & Lendínez, 1999; Carnero *et al.*, 2000; Cuetos-Vega *et al.*, 2007; Garcés *et al.*, 2004), and have been used in combination with neuroimaging techniques in the study of schizophrenia (e.g., Gilvarry *et al.*, 2001; Quan *et al.*, 2015), depression (e.g., Okada *et al.*, 2003), Parkinson's disease (Camicioli *et al.*, 1998) and bipolar disorders (Nishimura *et al.*, 2015). They have also been used in the study of both healthy and neurologically impaired bilinguals (e.g., Roberts and Le Dorze, 1997; Gollan *et al.*, 2002).

In attrition studies, it is a commonly performed task that is usually combined with other measures of lexical diversity, and these studies usually find that attriters perform lower than control groups (e.g., Schmid & Jarvis, 2014; Waas, 1996; Yağmur, 1997) or they perform better in their L2 than in their L1 (Ammerlaan, 1996). Other studies found no significant difference between groups in the performance of this task whereas they find significance in other fluency measures (e.g., Opitz, 2011). The problem with this type of task arises when the studies include factors that would predict within-group variation. Factors like educational level (Yağmur, 1997), and overall language proficiency (Schmid, 2006) were not found to correlate to VFT measures (Schmid & Köpke, 2009), whereas studies using other measures of lexical fluency did (e.g., Keijzer, 2007).

VFTs were used successfully in assessing the speech of adult aphasics (Goodglass *et al.*, 2001) and to assess the decrease of verbal fluency in healthy elderly speakers (see Goral, 2004). However, while these tasks do measure lexical diversity, they cannot be used alone as a tool to assess verbal retrieval difficulties. Therefore, they are not a valid instrument to assess adult non-pathological language loss, because they are performance tasks restricted to specific lexical domains (Köpke & Schmid, 2004), and attrition is a much wider phenomenon. Furthermore, VFTs cannot account for interference phenomena like code-switching or tip-of-the-tongue states, which do happen in spontaneous speech, and which some authors claim are phenomena relevant to the assessment of lexical attrition.

1.2.2 Picture naming and verification tasks

Another example of a technique used in early research is that of picture naming tasks (PNTs) and picture verification tasks (PVTs), used in some attrition studies (e.g., Ammerlaan, 1996; Hulsen, 2000; Kargar & Rezai, 2014; Tsimpli *et al.*, 2004). PNTs attempt to measure lexical access using visual stimuli: the participant is asked to either recall (picture *naming*) or recognize (picture *verification*) a series of pictures in any given language. In *naming* tasks, the participant is given a series of pictures, usually from the Snodgrass battery (Snodgrass & Vanderwart, 1980), and needs to name them in the language under study. This task is said to measure lexical recall or the active knowledge of vocabulary. In *verification* tasks, participants are given both the images and words and are instructed to match them in pairs. This task is said to measure lexical retrieval, or the passive knowledge of language. The results are usually measured combining reaction time (RT) with accuracy of the answers. Some authors use untimed tasks (Schoenmakers-Klein Gunnewick, 1998), while others measure time with a stopwatch (Isurin, 2000; Soesman, 1997). Hulsen (2000), on the other hand, used a more thorough approach including specialized software and hardware.

Both PNTs and PVTs, however, involve the activating of semantic information related to a certain visual representation, so it is not a strictly lexical process. To name a picture, the visual characteristics of the object need to be analyzed (and some authors claim that a visual representation must be activated), before activating the semantic (or conceptual) representation and activating then the phonological representation. The main findings in bilingual lexical retrieval and access studies are frequency and cognate effects, and can be explained using the Activation Threshold Hypothesis (Paradis, 2004): more frequent items are retrieved more easily, and not inhibiting a language may result in cognate effects. The same effects are found in attrition studies (e.g., Ammerlaan, 1996; Hulsen, 2000; Schoenmakers-Klein Gunnewiek, 1998).

Picture naming techniques were widely used in the early study of aphasia (e.g., Pease & Goodglass, 1978; Butterworth *et al.*, 1984), and have been proven useful in studies like Kohn and Goodglass (1985), who compared the performance of Broca's, Wernicke's and anomic aphasics and found that anomics committed less phonemic errors and more multiword circumlocutions, suggesting that they did not have difficulty producing words but accessing them. In the last decades it has been proven to be a useful technique when combined with neuroimaging techniques (e.g., Naeser *et al.*, 2005).

Other areas where picture naming and verification tasks are used are the study of linguistic deficits like dyslexia (e.g., Swan & Goswami, 1997) or SLI (e.g., Lahey & Edwards, 1999), and the study of language impairment caused by diseases like schizophrenia (e.g., Barrera *et al.*, 2005) or dementia (e.g., Garrard *et al.*, 2005).

1.2.3 Spontaneous speech

It has been argued that VFTs and PNTs should be combined with analyses of lexical diversity in (unguided) free speech (Schmid, 2004). By themselves, these measures do not cover the whole speaker's repertoire, and speaker performance in these tasks does not correspond to natural performance. Thus, they cannot be used alone to assess lexical attrition. For instance, some studies find no significant differences between the performance of attriters and controls in PNTs, whereas other diversity measures yield significant differences (e.g., Yilmaz & Schmid, 2012). Research on spontaneus speech is preferred because when speakers take part in these tasks they need to integrate information from all linguistic levels like in real-life situations. In these tasks, it could be that attrition effects would be even more pronounced, and these could be an indication of attrition being the result of incremental knowledge: the increased cognitive load involved in managing two language systems at the same time. Some trade-off effects might be observable, "for example between morphosyntactic complexity/accuracy and lexical diversity, or between the use of less frequent linguistic items and fluency" (Schmid & Jarvis, 2014: 731). Furthermore, experimental designs that focus on spontaneous speech can account for phenomena like code-switching, code-merging, and interference error.

This is why an increasing number of studies include elicited speech tasks, despite the difficulty of quantifying and interpreting free speech phenomena (Schmid, 2004). Elicited tasks usually consist of semi-structured interviews (e.g., Gross, 2004: Schmid & Jarvis, 2014; Yilmaz & Schmid, 2012), picture descriptions (Dostert, 2009), or the retelling of movie excerpts (e.g., Dostert, 2009; Keijzer, 2007; Opitz, 2011; Schmid & Jarvis, 2014). These interviews still face a problem, since it is uncertain how large the degree of freedom in the interviews is, and hence how naturalistic the obtained speech is. Ideally, measuring verbal fluency should aim for capturing the lexical diversity of the speaker in a completely spontaneous context, and this is better accomplished in interviews than in other elicited tasks. The more naturalistic the task is, the bigger are the differences that are usually found in attriters' and controls' lexical diversity (Schmid & Jarvis, 2014).

Analyzing free speech carries some methodological difficulties, mainly regarding the kind of analysis that is chosen. There are different tools to analyze lexical diversity, or measures of lexical diversity in free speech, and the use of some of them has been an object of debate in the literature. The main measures used are type and tokens, *vocd*-D, MTLD, or information-theoretic measures like Shannon's entropy index.

1.2.3.1 The type-token ratio and other related measures

The first approach to measuring the diversity in a text is to count the instances (*tokens*) of all the distinct words (*types*) that occur in the text. Some authors use the *type-token ratio* (TTR) to measure the arrangement of vocabulary in discourse. This measure, introduced by the speech pathologist Wendell Johnson (Johnson, 1939) is calculated using the total numbers of different words a speaker has used (types) and how many words he has used in total (tokens). This measure is not stable if we change sample length, which Johnson solved by using samples of the same number of tokens. This solution was not fully satisfactory (e.g., Malvern *et al.*, 2004). There were other attempts to find measures that overcame this sample-size dependency, one of them was Yule's K, a formula that was a sum of the probabilities calculated from the types occurring at each level of frequency (Jarvis, 2013), and represented the probability that, during the selection of random words from a text, the same word would be chosen consecutively twice (Baayen, 2001). Other measures like Guiraud's index or *vocd*-D relied on variations of Johnson's and Yule's formulations. Probably the most popular measure used in attrition studies is D, which is the result of the *vocd* calculation (cf. Malvern *et al.*, 2004) included in the CLAN software (MacWhinney, 2000a).

1.2.3.2 MSTTR, MTLD, MATTR and Shannon's entropy index

However, McCarthy and Jarvis demonstrated that length problems persisted even in measures like Guiraud's index, Yule's K and *vocd*-D (McCarthy & Jarvis, 2007). As Schmid & Jarvis (2014) explain, "the effects of length on probability-based measures such as Yule's K and *vocd*-D are relatively subtle, the principles of probability render them sensitive to a second factor: evenness" (Schmid & Jarvis, 2014: 731), which depends on "how evenly the tokens are distributed across types" (ibid.). Evenness is relevant to the study of lexical attrition because it can be an indicator of lexical access being contextually determined: when access to a specific item is compromised, it might be due to the activation of that item being determined by context. If the activation is spread across a smaller range of the semantic field, it would lead to a more uneven distribution of the tokens across the entire text than in the case of monolinguals (ibid.). This leaves us with two limitations this measure needs to overcome: on the one hand, it should account for textual evenness and, on the other hand, it should overcome the sample-size dependency problem.

The problem of sample-size dependency is solved by two measures. The first one was introduced by Johnson (1944): the *mean segmental type-token ratio* (MSTTR), a calculation of TTR from a standard number of tokens from each text, or alternatively, a calculation of the average TTR "over multiple, equally sized subsamples of a text" (Jarvis, 2013: 91). The second solution was the *measure of textual lexical diversity* (MTLD), introduced by McCarthy & Jarvis (2010, 2013). MTLD is calculated by holding TTR constant and calculating the average number of words in every segment of text that remains above the TTR cutoff value – usually 0.72 (Jarvis, 2013: 94). These two measures are mirrored, since they either keep word number or TTR constant and calculate the other variable.

An alternative measure similar to MSTTR and MTLD is the *moving average type-token ratio* (MATTR, Covington & McFall, 2010). This measure solves the type-token ratio text-length-dependency problem by using an algorithm that computes the TTR through a moving window (of constant tokens) that is independent of text length. This measure has proved useful for detecting both changes within a text when using bigger windows, as well as differences between texts when using smaller windows (id.).

The second problem, accounting for textual evenness, can be solved by applying the diversity measures that ecologists use to measure biodiversity. Diversity indices are measures that increase both when the number of types increases and when evenness (i.e., the distribution of tokens along types) increases. One of the preferred diversity indices is Shannon's informational entropy index (Shannon, 1948). This index represents how unpredictable are the results of a sampling process. Evenness is then calculated as a ratio of that index divided by the maximum possible diversity in the text if all types in the text were equally abundant.

But some authors (e.g., Jarvis, 2013; Schmid & Jarvis, 2014) claim that overcoming sample-size and evenness limitations is not enough to represent the diversity of a text. Jarvis (2013) claims that future measures of lexical diversity should not only be predictors of variables like proficiency rate but should also predict how humans would rate that diversity. As a starting point, studies of lexical diversity should include new measures for size (i.e., number of tokens), richness (i.e., number of types), effective number of types, evenness, disparity, importance and dispersion (Jarvis, 2013).

1.2.3.3 Findings from L1 attrition and other studies

Most attrition studies that look at vocabulary use calculate the number of types and the number of tokens. One of the most popular measures of lexical diversity is *vocd*-D, which is used in most L1 attrition studies (e.g., Keijzer, 2007; Opitz, 2011; Schmid & Dusseldorp, 2010; Yilmaz & Schmid, 2012) or early L2 attrition studies (e.g., Kang & Lee, 2013).

Schmid & Jarvis (2014) computed these and the other measures (evenness, disparity, importance and dispersion) on two populations of attriters (German L1 speakers in either a Dutch or English L2 environment) and a control population, and found that the applied measures that had a predictive power for group membership were those that measured the diversity and the distribution of types in the interview data, and not so much in the elicited narrative. In terms of distribution, a predictive factor for attrition was that attriters make less use of low-frequency vocabulary in elicited speech. This was also found by Yilmaz & Schmid (2012), who studied a group of Turkish-Dutch bilinguals and found similar results in frequency use, whereas they found no predictive power on controlled tasks.

In the study of aphasia, the use of tools like type count, TTR, or *vocd*-D is also quite extended (e.g., Fergadiotis & Wright, 2011; Wachal & Spreenn, 1973; Wright *et al.*, 2010). *vocd*-D has also been calculated in language acquisition studies (e.g., Durán *et al.*, 2004). Malvern and colleagues (2004) studied lexical diversity among very different groups: aphasics, acquirers or schizophrenics, using a variety of measures like type count, TTR, MSTTR, etc.

1.3 Zipf's Law

One of the first discoveries in vocabulary research is that there is a constant distribution of the word frequencies for every text, such that the rank of every word and its frequency are inversely related: the most frequent word is about twice as frequent as the second most frequent word, which in turn is about twice as frequent as the third most frequent word, and so forth. One of the first to study this phenomenon was George Zipf (Zipf, 1935), and the relationship was henceforth known as Zipf's law.

1.3.1 Formulation

Zipf's formulation of the law was that, if words were to be ranked according to frequency, for the rth most frequent word w with a frequency of $f(w)$ the relation

$$f(w) = \frac{C}{r(w)^{\alpha}} \tag{1.1}$$

held for an $\alpha \approx 1$ (Zipf, 1936, 1949). In the equation, $r(w)$ would be the *frequency rank* of a word, and $f(r)$ its frequency in a natural corpus. C is a constant that is determined by the size of the text. From the formulation it follows that the most frequent word ($r = 1$) has a frequency equal to C, the second most frequent word ($r = 2$) has a frequency of $\frac{C}{2^{\alpha}}$, etc.

The Zipfian distribution (1.1) is a power law that as such has the property of becoming a linear function when taking the logarithm of both sides, giving:

$$log f(w) = log C - \alpha log r(w). \tag{1.2}$$

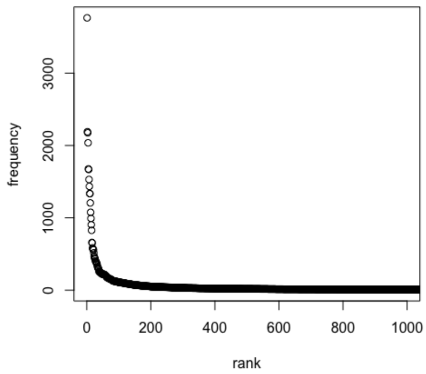
In the obtained linear function (1.2) the exponential α in Formulation 1.1 becomes the slope, and the logarithm of the constant C becomes the intercept. These values can be easily calculated using maximum likelihood estimation (Izsak, 2006, for R implementation see Murphy, 2015).

An example of how this distribution looks can be seen in Figure 1.1, where I calculated the word frequency distribution of Oscar Wilde's *The Picture of Dorian Gray* (Wilde, 1994). As we can observe in Figure 1.1a, the words from the higher ranks (and lowest frequencies) appear to accumulate and form a straight line in the end, meaning that the words with the higher ranks have all the same frequency, which is just one occurrence. These occurrences are known as hapax legomena, and in corpus linguistics they are defined as the items that occur just once in the text. In the side of the plot of lower ranks (and higher frequencies) we observe a similar phenomenon, where the slope seems to be less inclined. This is observed in large corpus data, and it is due to the fact that "high frequencies tend to be lower than what would be predicted by their rank according to Zipf's law" (Baroni, 2009: 14). In Figure 1.1b we can see how high and low rank values make the slope deviate from the ideal linear distribution in the logarithmic formulation (1.2), making α diverge from the actual slope of the distribution. This is why Zipf's law was corrected by Mandelbrot (1953), who added a parameter to take care of this downward curvature, giving Formulation 1.3:

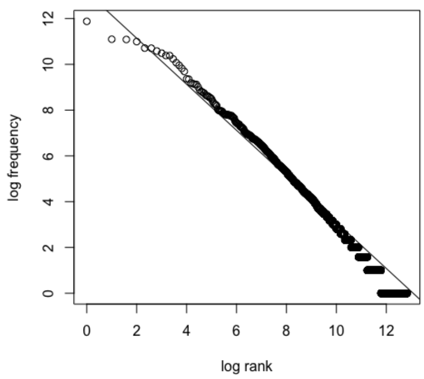
$$f(w) = \frac{C}{(r(w) + \beta)^{\alpha}}. \tag{1.3}$$

This new formulation (1.3), known as Zipf-Mandelbrot's law, includes Zipf's law as the instance in which the component $\beta = 0$. Usually, we will expect the values $\alpha \approx 1$ and $\beta \approx 2.7$ (Mandelbrot, 1953, 1961; Zipf, 1936, 1949), but when the latter is reasonably small, it will lower the frequency of the first few ranks significantly, but will not affect higher ranks. If we take logarithms in both sides of the formula, the resulting function will no longer be a straight line, i.e.,

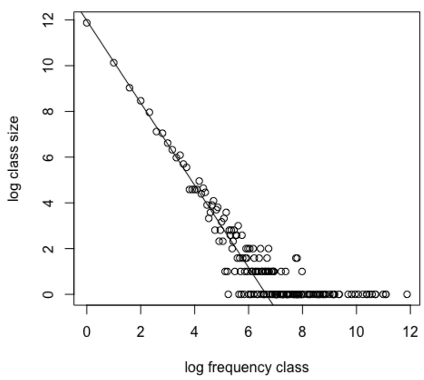
$$log f(w) = log C - \alpha log(r(w) - \beta). \tag{1.4}$$



(a) Word frequency - word rank plot of first 1000 tokens (Zipfian distribution)



(b) Log rank - log frequency plot of all tokens (alpha distribution)



(c) Log class size - log frequency class plot (beta distribution)

Figure 1.1: Word frequency distributions in Oscar Wilde's *The Picture of Dorian Gray* (Wilde, 1994)

From Formulation 1.3 we can also derive the *alpha* and *beta* formulations. If $f(r)$ is the frequency of the rth word, then we have that

$$f(r) \propto r^{\alpha}, \tag{1.5}$$

with usually $\alpha \approx 1$. This formulation is known as the *alpha formulation* of Zipf's law, and like Formulation 1.3, it corresponds to the distribution in Figure 1.1b. They are basically the same formulation, but leaving out the constant C from Formulation 1.3. Another possible formulation of Zipfian distributions is the *beta formulation*. Unlike Formulations 1.3 and 1.5, this formulation is not based in word ranks, but is based on frequency classes, i.e., how many words occur once, how many words occur twice, and so forth. If $P(f)$ is the proportion of words whose frequency is f in a text, we obtain that

$$P(f) \sim f^{-\beta}, \tag{1.6}$$

where we usually get a $\beta \approx 2$. In other words, the beta formulation tells us that the number of words with frequency f is approximately equal to the frequency f to the power of $-\beta$. The β exponent in the beta formulation (1.6) should not be confused with the β exponent in Zipf-Mandelbrot's formulation (1.3), since the latter is the result of adjusting Formulation 1.1 to accomplish a better fit of the first few ranks, whereas the former is the result of a new formulation of Zipf-Mandelbrot's formulation (1.3) using frequency classes. This formulation allows for the analysis of word frequencies without needing word rank, leading to a distribution that corresponds with that in Figure 1.1c, where class size is plotted on frequency class size ($P(f) = 1, 2, 3...$). In the case of *The Picture of Dorian Gray* (Wilde, 1994) we can clearly see that the value of the slope β comes close to -2 (specifically -1.8). The problem with this formulation is that, in large texts, this distribution is only accurate for the frequency classes of smaller frequencies, so it is not a useful tool for smaller texts. In Section 1.3.3. I will discuss these values and its possible relevance for the study of lexical diversity or linguistic complexity.

1.3.2 The origins of Zipf's law

What makes Zipf's law intriguing is that it is a distribution that happens in all sorts of natural language, regardless of the form (written or spoken) or the language in question. Its universality has even been assessed in extinct languages or languages that have not been translated yet, like Meroitic (Smith, 2008). The question of where Zipf's law originates is quite controversial and there is still no agreement on its actual origin. There are diverse explanations for the law, but these can be grouped in two opposing views: those who think that Zipf's law originates from statistical processes, and those who think that it originates from the functioning of psychological processes. Following Piantadosi (2014), statistical theories would set the origin of Zipf's law in random concatenative processes (e.g., Conrad & Mitzenmacher, 2004; Li, 1992), in mixtures of exponential distributions or the invariance of power laws under aggregation (Farmer & Geanakoplos, 2006), scale-invariance (Chater & Brown, 1999), optimization of entropy (Mandelbrot, 1953), multiplicative stochastic processes (Mitzenmacher, 2004) or random walks on logarithmic scales (Kawamura & Hatano, 2002). On the other hand, psychological accounts would set the origin of Zipf's law in semantic organization (van Egmond *et al.*, 2015; Guiraud, 1971; Manin, 2008), in communicative optimization (Ferrer i Cancho, 2005a, 2005b; Ferrer i Cancho & Solé, 2003; Zipf, 1936, 1949) or optimized memory search (Parker-Rhodes & Joyce, 1956).

However, Zipf's law is not a law that is exclusive to natural language, but it also appears in other human systems like music (e.g., Manaris *et al.*, 2005), computer systems and hardware instructions for programming languages (e.g., Chen, 1991), levels of abstraction in software (Louridas *et al.*, 2008), and many aspects of the internet (Adamic & Huberman, 2002). It occurs also in the distribution of a large number of physical and biological systems (e.g., Farmer & Geanakoplos, 2008; Li, 2002; Mitzenmacher, 2004). So this complicates making holistic theories that account for all these diverse phenomena. An interesting view is that of Piantadosi (2014), who claims that none of the models of Zipf's law has explanatory power to account for Zipf's law in word frequencies, but believes that the lack of a unitary explanation is probably due to the fact that there are multiple causal forces at play.

In the following sections I will address three different accounts of Zipf's law. The first holds that Zipf's law originates in probabilistic models, being a statistical account of the law. I will oppose this view to two other views that justify Zipf's law in the functioning of psychological processes. The first view is that of Zipf (1936, 1949), and postulates that the law is a result of communicative optimization. Then, I will explain a different view that holds that the origin of the law is to be found in the organization of human memory.

1.3.2.1 Zipf's law originates from probabilistic models

Zipf's law is a ubiquitous law that appears in many structures other than word frequency distributions. This is why many authors believe that the law is essentially a result of statistical systems and its existence does not contribute any further meaning about language structure. There are two (main) different probabilistic accounts for Zipfian distributions: random-typing accounts and stochastic models. A very common anecdote used to illustrate the first is to hypothesize that if a monkey randomly typed on a typewriter, it would randomly type words (because he would also press the spacebar) and the randomly generated text would also follow a Zipfian distribution (Conrad & Mitzenmacher, 2004; Li, 1992; Miller, 1957). This theory is also known as intermittent silence, since it relies on the idea that the probability that the monkey would press the spacebar is the same as any other key on the keyboard. As a result, short words would be ranked as the most frequent and long words would be the least frequent. To test this hypothesis, Piantadosi (2014) analyzed the word frequency distribution of a corpus where the word boundaries were not delimited by a space character, but by the letter "e", and found a near-Zipfian distribution, but with deviations from the shape of the curve of natural language.

The difference with natural language is that intermittent silence theories would predict that the frequency of a word is, in a way, determined by its length in number of characters, and, in real language, this relationship does not exist. In this sense, random-typing accounts do not predict natural language frequency distributions. Another limitation is that random monkeys theories obviate that language is not generated at random, since words are selected from a repertoire of existing items, and are actively selected in order to convey a particular meaning. In this regard, random-typing accounts are poor theories of language production. There are also distributions of frequency within language structures that this theory does not account for; for instance, it cannot account for the existing frequency differences between word categories (i.e., content words are less frequent than functional words), or why there is a certain systematicity in the word frequency of certain semantic categories like the ones found in numerals ("one" is more frequent than "two", which is more frequent than "three", etc.) This theory cannot account for the variance in the exponent of the law, as will be discussed later in this chapter.

The second statistical account of Zipf's law posits the existence of stochastic models which are determined by word usage. In probability theory, stochastic models, as opposed to deterministic models, explain the evolution of a system with a collection of random variables. These systems are characterized by some indeterminacy: even if the initial state is known, there are several possible directions in which the system might evolve. Stochastic accounts of Zipf's law rely on the idea of "preferential reuse," which is how words tend to be reused in a text. Introducing a word in discourse makes it more likely for that word to appear again later in discourse. They posit that this will lead to "a very skewed frequency distribution, since frequent words will tend to get reused even more" (Piantadosi, 2014: 15). These models usually assume that there is a constant probability of introducing a new word combined with the preferential reuse of previous words, which is a random variable. Examples of these models can be found in Mitzenmacher (2004) or Baayen (2001).

These models generate words at random but with the constraints of a certain reuse of frequent words and a constant probability of introducing new words. Despite getting closer to natural language generation, they do not serve as a theoretical framework to explain Zipf's law in natural language because they do not capture the intentional aspect of meaningful language production. Another limitation of these models is that they do not explain why some words are more frequent than others, particularly function words: if word frequencies were generated by chance, the most frequent words would not necessarily have to be function words. The relevance of these theories is that they show that if language generation followed stochastic models, then they would give rise to Zipf's law. However, they do not account for the particularities of language that cannot be captured by random models, as mentioned in the previous paragraph.

It is believed that these models may be useful in explaining some simple situations of language use. For instance, Piantadosi (2014) shows how these models can explain the near-Zipfian use of novel words, and argues that simple stochastic systems might be the underlying force of the law, but constrained by other factors like syntax. As he posits, the law probably originates from a confluence of these models with other factors like syntax or discourse that are not simply explainable with preferential reuse: "words are reused in language [...] not because of an intrinsic preference for reuse itself, but instead because there is a latent hidden variable, a topic, that influences word frequencies" (Piantadosi, 2014: 16).

1.3.2.2 Communicative optimization

Communicative accounts (e.g., Ferrer i Cancho, 2005a, 2005b; Ferrer i Cancho & Solé, 2003; Zipf, 1936, 1949) claim that the origin of Zipf's law is found in the least effort principle: if a communicative repertoire is too unified or repetitive, then only a few messages are made available to convey a big battery of common items, and then communicative complexity will be low. In these accounts, the law arises from communicative optimization principles. What Zipf's law does is to maximize this efficiency and to minimize the communication costs (Ferrer i Cancho & Solé, 2003). Grammar and lexicon are emerging aspects of language linked to the connectivity between linguistic elements (Bates & Goodman, 1999), and as a result high frequency words are usually words with little semantic content and enormous syntactic importance, since they are strongly connected in the lexical network (Ferrer i Cancho, 2006), whereas low frequency words are usually less relevant syntactically and of higher semantic content (Ferrer i Cancho *et al.*, 2005). Numerically, this would result in a power law where the higher frequency elements are the more communicatively efficient (and then used proportionally more often) and the low frequency words would be sorted by communicative relevance.

A particular assumption of Ferrer i Cancho and colleagues (e.g., Ferrer i Cancho & Solé, 2003) is that meanings are equally likely to be conveyed, an assumption that is questioned by Piantadosi (2014). Later attempts leave out this assumption, but still do not incorporate known factors like memory latency, frequency effects or context-based expectations (id.). In their framework, speaker effort is proportional to the diversity of signals they must use, and the listeners pay a cost that is proportional to the entropy over referents given a word. They use a parameter which measures the trade off of the costs between speaker and listeners, which at a particular value ($\lambda = 0.41$) recovers a Zipfian distribution. When the parameter comes closer to $\lambda = 0.5$ or $\lambda = 0.3$ Zipf's law cannot be recovered, which has been argued to be a non-robust aspect of the theory.

Piantadosi (2014) also argues that these accounts could, with difficulty, explain near-Zipfian distributions arising from fixed referential content: he notes that there are domains of objects whose reference is highly constrained by the natural word (like months, planets, or elements) that normally follow near-Zipfian distributions, and it is not clear why communicative complexity would constrain such referentially fixed content. On the other hand, the theory cannot explain the presence of the law in other human systems, or the presence of the law in the learning of novel words (explained in next section).

1.3.2.3 Zipf's law originates from memory structure

Some authors claim that Zipf's law originates in semantics: frequency of words might be determined by the meaning of words, and semantic organization may be giving rise to this frequency distribution. Manin (2008) argued that these frequency distributions are the result from the labeling of semantic hierarchies combined with avoiding using synonyms. However, Zipfian distributions are also found in words whose referent is constrained by the natural world, such as months, planets, or elements; so the choice of semantic referent by the lexicon does not explain these frequency distributions (Piantadosi, 2014). Semantic explanations must yet be developed to fully account for these phenomena.

The origin of these distributions might be behind the psychological processes involving word production, and some authors claim that its origins reside in the structure of the lexicon or, more broadly, the structure of human memory. Human memory, or remembering, has been characterized as following power laws modulated by time of learning (e.g., Wickelgren, 1977; Wixted, 2004a, 2004b; Wixted & Ebessen, 1997), and these power laws are observable in the scaling relationships observed in other aspects of cognition (Kello *et al.*, 2010). If these scaling properties of memory were the real origin of Zipfian distributions, "it could provide a parsimonious and general explanation, able to unify word frequencies with memory, while also explaining the occurrence of related laws in other systems humans use, such as computer software and music" (Piantadosi, 2014). This theory, however, would still need answer the question of why human memory is structured in a way that its output conforms to Zipfian distributions.

Piantadosi (2014) found that power laws appeared when participants were presented with novel words. In the study, participants were presented with the names of eight different characters of a story that participants had to continue in at least 2000 words. The relative frequency of these names was found to conform to a power law, with individual differences regarding which name ranked first or second. This could be illustrated as an example of Zipf's law being the result of the structuring of (new) information. In this line of work, Steyvers and Tenenbaum (2005) justify this organization in how the lexicon is organized, claiming that in semantic structures the number of connections between words follow a power law. It could follow that the retrieval of these would automatically result in Zipf's law.

More evidence for Zipf's law originating in the mental lexicon comes from the study of aphasia. Van Egmond and colleagues (van Egmond, 2011; van Egmond *et al.*, 2015) studied the spontaneous speech of aphasic patients and hypothesized that the origin of Zipf's law was either in the processing of words or in the storage of words. Aphasic patients are know for presenting difficulties in word retrieval, and it is believed that these difficulties are due to processing deficits and not to a loss of knowledge (Avrutin, 2006; van Ewijk, 2013). They found that aphasic patients' spontaneous speech conformed to Zipfian distributions, suggesting that the origin of Zipf's law is the mental lexicon, and not the word selection and retrieval mechanisms.

1.3.3 The exponential component

The exponential component of Zipf's law has received some attention in linguistic studies. While the usual value is $\alpha \approx 1$ for Equation 1.5 or $\beta \approx 2$ in Equation 1.6 (Ferrer i Cancho, 2005a; Zipf, 1942), some studies found significant deviations from these values. When referring to the exponential component, each study uses either α or β depending on the formulation of choice of the study, but these exponents are proportional (Chitashvili & Baayen, 1993) given that Equations 1.5 and 1.6 are equivalent with

$$\beta = \frac{1}{\alpha} + 1. \tag{1.7}$$

It should be reminded that this β corresponds to the exponent in the β formulation of Zipf's law (1.6), not to the β in Zipf-Mandelbrot's formulation (1.3). Another confusing aspect of the literature about referring to the exponents is whether they are considered to include the negative sign or are referred to with their absolute value (ignoring their negativity). In this study I will refer to these exponents with their absolute value, so when talking about increases or decreases in the value they should be interpreted as increases or decreases in the absolute value of these, despite their negativity. In the next sections, the different findings relating to the exponent of the laws are presented, and then discussed according to the current theoretical frameworks in use.

1.3.3.1 Previous findings

In patients with fragmented discourse schizophrenia, α values lower than 1 (β values higher than 2) have been found (Piotrowska *et al.*, 2004; Piotrowski *et al.*, 1994). The speech of these patients is usually characterized as very chaotic and varied, lacking a consistent subject. In patients with advanced forms of schizophrenia, whose speech is characterized by the repetition of words and word combinations relating to the patient's obsessional topic, the α values found are bigger than 1, with β values usually in the range $1 < \beta < 2$ (Ferrer i Cancho, 2005a; Piotrowska & Piotrowska, 2004; Piotrowski *et al.*, 1994).

In the study of child speech, α values higher than 1 ($\beta < 2$) have been found (Ferrer i Cancho, 2005a; Hernández-Fernández, 2005). Piotrowski and colleagues (Piotrowski *et al.*, 1995; Piotrowski & Spivak, 2007) found that values of α clearly exceeded 1 in children. Baixeries *et al.* (2013) analyzed the evolution of the exponential component in children in a longitudinal study and found that the exponent α tends to decrease over time, a tendency that was not found in control adults. They also measured the mean length of utterances (a measure of syntactic complexity) and found that this increased as the exponent decreased, suggesting that the exponent of Zipf's law and linguistic complexity are interrelated.

In the study of other linguistic pathologies, values of the exponent have been reported for aphasia and Alzheimer's disease. Van Egmond *et al.* (2015) found that aphasic speakers with anomia had a significantly higher α than the control group. Anomics' difficulties with word retrieval have been seen as the result of processing deficits, so this difference in slope from that of controls might be a reflection of these deficits. Hernández-Fernández and Diéguez-Vide (2013) also found that people with Alzheimer's disease showed a significantly different β than the control group. Specifically, they found a higher βin the Alzheimer's patients group, which is an unexpected result. They argue that their results are still meaningful despite the unexpected value, because they show that there is a difference between groups. However, an examination of their methodology reveals serious deficiencies, since they did not compare texts with the same size, and the exponents are known to be highly dependent on sample size.

1.3.3.2 The meaningfulness of the exponent

There are some theoretical models that have shown that various aspects of the complexity of a communicative system, like its capacity to combine words to build more complex sentences, may depend on the value of the exponent (Ferrer i Cancho, 2006; Ferrer i Cancho *et al.*, 2005). These models claim that the variations in the values of the exponent of the law rule out theories that neglect the meaningfulness of Zipf's law, like random accounts. Intermittent silence theories do not predict the exponents found in human language, but a different range of exponents, and they also conceive speech generation as a sequence of uncorrelated words, while it is widely known that, in human language, there are long-distance correlations among text elements.

In communicative optimization theories, Zipf's law is still seen as the result of the least effort principle (Zipf, 1942). They argue that the human ability to successfully communicate maximizing the information transfer at the cost of communication is observable in the exponent. They rely on models that see word frequency as an epiphenomenon of word meaning to account for the variation in the exponent, i.e., that the growth of β or the decrease in α is associated with a greater semantic precision. Applying an information theoretic paradigm, the exponent contains information regarding the balance between cost of communication and communicative efficiency: increases in β (or decreases in α) are argued to correlate to increased efficiency (Ferrer i Cancho, 2005a).

In these models, the exponent is also explained in terms of neural connectedness. In this scenario (Ferrer i Cancho, 2006), they model communication as systems where signals are associated with stimuli, and assume that the signals already follow Zipfian distributions. In this case, the exponent still depends on a balance between maximizing the information transfer and saving the cost of signal use. The variation in the exponent is seen as a result of the connectedness of the system: increased connectedness would result in an increased possibility of starting from a signal (or stimulus) and reaching the remaining signals and stimuli of the network, which would result in change in the exponent (specifically, a higher β). This connectedness is tightly related to syntax and symbolic reference (Ferrer i Cancho *et al.*, 2005), but can also be related to synaptic density: a higher value of the exponent β (or a smaller α) would correspond to an increased synaptic density, that is, increased connectedness. Decreased synaptic density is found in patients like schizophrenics in the acute phase, who show smaller values of β than controls (Ferrer i Cancho, 2006). As described before, the speech of these patients is characterized by the repetition of words and word combinations relating to the patient's obsessional topic.

This notion of connectedness could account for the increases in the exponent in child speech found in the longitudinal study by Baixeries and colleagues (Baixeries *et al.*, 2013). Children show a progressive reduction in the exponent that could be explained in terms of development of connectedness: the emergence of syntactic communication could be "a phase transition to connectedness in the network of word syntactic interactions" (Ferrer i Cancho, 2006: 251). Baixeries and colleagues provided empirical evidence of the exponent being related to the mean length of utterances, a measure of syntactic complexity, but this could be an epiphenomenon of overall connectedness. This notion of connectedness could also explain the results from van Egmond and colleagues (van Egmond *et al.*, 2015), who found that aphasic speakers with anomia had a significantly higher α and a significantly lower β than the control group. The difference in performance between anomics and controls could be explained by the reduced connectedness in the aphasics' linguistic systems, altered after stroke, and would predict that they would follow a similar pattern to acquirers during speech therapy.

A possible answer to why Zipf's law exists is that given by universality theories. It could be that memory mirrors sensorial input, and Zipf's law originates from the distributions that are observed in the world. This assumption is usually contained in communicative optimization accounts. For instance, Corominas-Mutra and Solé (2010) derived Zipfian distributions from algorithmic information theory. Later, they hypothesized that Zipf's law in word frequency distribution is the result of sample-space reduction of word usage along sentence formation (e.g., Corominas-Mutra *et al.*, 2015; Thurner *et al.*, 2015). They explain the exponent $\alpha \approx 1$ with the notion of nestedness, a similar concept to that of semantic precision. When a word appears in discourse, it restricts the possibilities of the words that can follow, both at the grammatical and the semantic/contextual level. A text shows more nestedness when its words are more restricted. They claim that the exponent alpha is related to this nestedness in an approximately linear way, and is also related to text length (Font-Clos *et al.*, 2013; Thurner *et al.*, 2015). To my knowledge, however, this framework does not yet account for the variation in the exponent in the groups that has been covered here.

So even if the linguistic relevance of the exponent is not fully clear, it is likely that it correlates to textual/semantic/syntactic complexity, or to the entropy, or the amount of information transmitted by the text. From the definition of the law, it follows that the exponent relates to the amount of more frequent and less frequent items: when the exponent α is higher, the number of high frequency items increases (and the number of low frequency items decreases), and the same happens reversely. What is certain is that the idea of Zipf's law being a power law with a constant exponent needs to be abandoned, and the next questions to be answered should be, in the first place, why human language presents a particular range of values for the exponent, and, in the second place, why there is variation among human groups at all. Intuitively, previous research shows that deviations from the ideal $\alpha \approx 1$ can be linked to lexical richness: higher values of the exponent are found in children (Baixeries *et al.*, 2013; Piotrowski *et al.*, 1995; Piotrowski & Spivak, 2007) or anomic aphasics (van Egmond *et al.*, 2015). From the formulation of the law it also follows that a higher alpha means that the proportion of words from high ranks is higher than normal in comparison to words of lower ranks, meaning that the text has more frequent items and fewer non-frequent items. The relation between this value and other measures of linguistic performance still needs to be specified.

Chapter 2

Methods

2.1 Research Questions

In the previous sections I provided an overview on the literature regarding language attrition in general, and lexical attrition in particular. Attriter speech has often been described as having a smaller amount of lexical items, and their frequency distributions usually consist on highly frequent words being used more often, and low frequency items being less common. The study of lexical attrition needs to rely on the different techniques used in lexical statistics that have proven to be of different effectiveness when applied to attrition, acquisition or other studies.

In the present study I will use the data collected by Keijzer (2007). She studied the regression hypothesis testing a group of Dutch migrants living in Anglophone Canada. She compared the group's performance with a control group of Dutch monolinguals and a group of language acquirers living in the Netherlands. Among a variety of tests, general proficiency was assessed by means of a C-test that threw that controls performed significantly better than attriters, which in turn performed significantly better than the acquirers. She also assessed lexical diversity calculating *vocd*-D, and found the same relation of performance between groups.

To my knowledge, there is no study using Zipfian distributions in L1 attrition data. Using the data in Keijzer (2007) will allow me to address the follow research questions:

1. Does attriter speech conform to Zipf's law?

2. If (1) holds, is it with the same or a different slope to controls (and acquirers)?

3. Is the exponent in Zipf's law a strong predictor for group membership when compared to other measures of lexical diversity?

The lack of Zipf's law studies on L1 attrition (or bilingual data, more generally) makes question (1) remain unanswered, although there is no theoretical reason to believe that bilingual speakers would show a different word frequency distribution than that of monolingual speakers. Here I will attempt to confirm (1) and analyze the slopes of the distributions in the different experimental groups in order to give answer to question (2). I will do so in an attempt to weigh the predictive power of Zipf's law regarding lexical diversity. For that, the data from Keijzer (2007) will be analyzed according to the methodology in van Egmond and colleagues (van

Egmond *et al.*, 2015), who used these distributions to analyze the differences between aphasic and healthy speech. This will open the possibility of comparing the findings in attriter speech with other findings in studies that apply Zipf's law to speech pathologies such as aphasia or dementia.

Furthermore, I will include a comparison of lexical diversity measures (i.e., TTR, TTR for same size texts, MATTR, and VOCD), and I will compare their predictive power for the different experimental groups, in order to answer question (3). The results in this and previous studies will be later discussed focusing in three main points: the usefulness of the exponent in Zipf's law as a measure of lexical diversity, the meaningfulness of the law itself and its significance regarding (lexical) attrition and speech pathologies, and the organization of the mental lexicon.

2.2 Participants

As already metioned, Keijzer (2007) studied the regression hypothesis testing a group of Dutch migrants living in Anglophone Canada. She compared the group's performance with a control group of Dutch monolingual speakers and a group of language acquirers living in the Netherlands. The different experimental groups will be henceforth referred to as the attriter, the control, and the acquirer groups, respectively. The first two (attriters and controls) consisted on a total of 45 participants per group, while the acquirer group included 35 participants.

The attriter group was formed by first generation Dutch migrants that were interviewed in Ontario, in Anglophone Canada. Dutch migrants in Canada are known as being historically welcome because of their easy adaptability to the language and culture of the country (Keijzer, 2007: 144). The participants were chosen based on particular criteria. One criterion was that they should be older than 15 when they migrated, but had lived in Canada for longer than 20 years at the time of the study. This helped to control for incomplete acquisition (see Section 1.1.2). On the other hand, they had to be married to Canadian spouses in order to control for cases of high contact with the L1. This also helped to control for cases of language reversion (see Section 1.1.4).

The control group consisted of Dutch native speakers that had not spent a considerable time outside the Netherlands. They were matched with control groups on a one-to-one basis, and since attriters were tried to be matched with controls as similar as possible, these would not only be of similar ages and same gender, but were often siblings or relatives.

The acquirer group consisted on 35 secondary school students that were included in the study for the purpose of studying the regression hypothesis. The main motivation was that if there were mirror symmetries to be found between language attrition and language acquisition, these would be observable in the later stages of acquisition rather than in the early ones, given that language attrition is manifested as a very slow process. I decided to include this group in the study to see if the results corroborate the findings by Baixeries *et al.* (2013), who analyzed child speech in a longitudinal study and found that the Zipfian exponent α tends to decrease over time, tendency that was not found in adult controls.

From the 125 participants (45 attriters, 45 controls and 35 acquirers) included in Keijzer's study, the data of a total of 117 participants were included in the present study. There were two reasons for excluding the other eight participants. Firstly, the data of 3 participants (one from the control group and two from the attriters group) could not be retrieved. On the other hand, the data from the other five participants were excluded because of quantitative reasons: the study material used in this study consists on one of the various tasks performed by Keijzer (2007), and for these participants the content of this particular task (number of tokens) was considerably low when compared to the other participants. A decision was made to prioritize length of study material over number of participants, but trying not to alter the distribution of participants significantly. This left out of the analysis a total of one attriter, one control, and three acquirers, leaving a new group distribution of 42 attriters, 43 controls and 32 acquirers.

2.2.1 Sociolinguistic variables in participants sample

Keijzer (id.) included the following extralinguistic variables: age, gender, educational level and region of birth. Except for participant age, these were included in the present study both for the sake of continuity with Keijzer's work and to analyze the predictive power of the different measures of lexical diversity. Intuitively, if we assume that these sociolinguistic variables affect lexical diversity, it follows that if certain measure predicts more differences between groups, it can be said to perform better than measures that predict less differences. Therefore, sociolinguistic variables can be included to assess the performance of the different lexical diversity measures, and can be used to compare these to the eventual performance of the exponent in Zipf's law.

2.2.1.1 Age

As mentioned in Section 1.1.3, aging can have an effect in healthy monolingual speakers who are not in language contact environments. However, in this study, participants from the attriter and control groups were matched in age, and their age range does not show much variance (cf. Table 2.1). The differences in this variable might be still observable in the acquirer group. However, information regarding the age of participants was not complete, so this variable was not included in the analysis.

An overview of the distribution of the age of the participants included in the study can be seen in Table 2.1.

group	mean age	SD	age range
attriters (n = 42)	66.40	7.53	41-79
controls (n = 43)	66.20	8.17	45-80
acquirers (n = 32)	13.91	0.67	13-16

Table 2.1: Age distribution of participants

2.2.1.2 Gender

Previous findings in linguistic studies suggest that girls show higher aptitude for language learning than boys, and gender has been found to be an important predictor in attrition studies as well (Keijzer, 2007; Schmid, 2002). Effects of this variable on lexical diversity might be observable. An overview of the participants' gender can be seen in Table 2.2.

group	male	female
attriters (n = 42)	19 (45.2%)	23 (54.8%)
controls (n = 43)	20 (46.5%)	23 (53.5%)
acquirers (n = 32)	18 (56.2%)	14 (43.8%)

Table 2.2: Distribution of the gender of participants among groups

2.2.1.3 Educational level

As explained in Section 1.1.4.3, Dynamic Systems Theory claims that there is an interplay between system-internal (like intelligence, linguistic capacity, time or motivation) and system-external resources (like language and social input). The educational level of a person is a reflection of the confluence of both these system-internal and system-external factors, so it might have an impact on the speakers' lexical diversity and the degree of attrition.

	attriters (n = 42)	controls (n = 43)	acquirers (n = 32)
primary school	3 (7.1%)	1 (2.3%)	
secondary school basic	17 (40.5%)	14 (32.6%)	22(68.8%)
secondary school plus	8 (19.0%)	10 (23.3%)	10 (31.2%)
college	6 (14.3%)	10 (23.3%)	
university	8 (19.0%)	8 (18.6%)	

Table 2.3: Distribution of the educational level of participants among groups

In Table 2.3 a relation of the participants' educational level is provided. The classification of the academic background in different levels is done according to the Dutch educational system. In this system, after finishing their primary education, students are oriented towards a basic secondary education (named VMBO) or a selective secondary education (either HAVO or VWO). This selective secondary education prepares students to either access higher education at a applied sciences level (known as HBO, and referred as *college* in Table 2.3) or at a research level (known as WO, and referred to as *university* in Table 2.3).

In Table 2.3, participants are ranked according to the higher education level that they obtained, with the exception of the acquirer group, who are included in the education level that they were studying at the time that the experiment was conducted.

2.2.1.4 Region of birth and upbringing

Although it would not be expected to find any effect of the region of birth and upbringing of participants, I included this variable for the sake of continuity with Keijzer's study (Keijzer, 2007). She used a division of the Netherlands with three main regions: North, including the provinces of Friesland, Groningen and Drenthe; Central, comprising the provinces of Overijssel, Gelderland, Utrecht, Flevoland, North Holland and South Holland; and South, including the provinces of Limburg,

Figure 2.1: Three regions of the Netherlands (north, center and south) used in the study. Retrieved from Keijzer (2007: 253).

Brabant and Zeeland. A graphical representation of these areas can be seen in Figure 2.1.

	north	central	south
attriters (n = 42)	16 (38.1%)	16 (38.1%)	8 (19.0%)
controls (n = 43)	5 (11.1%)	28 (65.1%)	9 (20.9%)
acquirers (n = 32)		32 (100%)	

Table 2.4: Distribution of the region of origin of participants among groups

Table 2.4 shows the number of participants from each group according to their region of origin. Three of the participants (2 attriters and 1 control) were not included in the table, since they were not born in the Netherlands, but in Indonesia, which used to be a Dutch colony at the time the participants were born. These participants, however, moved to the Netherlands during their early childhood (Keijzer, 2007: 153).

2.2.2 Language proficiency of participants

Language proficiency of participants was assessed in Dutch for all participants, and also in English for the attriter group. This was done by means of a cloze test (C-test), where participants were presented with a set of different texts where some items (either content or function words) were missing. Prior to providing this texts to the participants, they were first administered to native speakers, and only texts where native speakers scored between 87 and 90% of correct answers were included in the experiment.

In the Dutch C-test, Keijzer found that controls outperformed attriters, who themselves outperformed acquirers. These differences between groups were found significant with a statistical analysis of variance and a Games-Howell post-hoc procedure (id.: 184). She also included the sociolinguistic variables in the analysis and found that age, gender and region of birth had no impact in the outcome of the Dutch C-test, but education level did: highly educated participants tended to produce higher scores (id.: 187).

In the English C-test, she found that attriters showed better performance in the English C-test than in the Dutch C-test (id.: 190), and again, there was an effect of educational level but no effect from other sociolinguistic variables (id.: 191).

2.3 Materials

It has been argued that language attrition studies should make use of natural language situations where all linguistic levels have to be integrated in real time (Schmid & Jarvis, 2014), because if there are attrition effects to be found, these would be more pronounced in such situations. This is why an increasing number of studies include elicited speech tasks or interviews. In her study, Keijzer (id.) included a film-retelling task, which has also been used in other attrition studies (e.g., Cherciov, 2011; Dostert, 2009; Opitz, 2011; Schmid & Dusseldorp, 2010; Schmid & Jarvis, 2014).

The task consisted in retelling an excerpt of Charlie Chaplin's *Modern Times* (for a scene-by-scene description of the excerpt see Keijzer, 2007: 353-355). The excerpt consisted of 20 minutes of film that were presented to the participants, who were later asked to retell what they remembered from what they had seen. The participants' answers were recorded and later transcribed in the CHAT transcription format (MacWhinney, 2000a).

2.4 Procedure

Two different analyses were performed. First, I used the transcripts from the film retelling task to assess lexical diversity. The following measures of were calculated: number of tokens, number of types, type-token ratio (both sample-size dependent and independent), *vocd*-D and MATTR. Later on, the same transcripts were used to study Zipf's law. These two analyses will be explained in the following sections.

2.4.1 Measuring lexical diversity

Types and tokens cannot be used as measures of lexical diversity by themselves since they are highly dependent on text size. However, the number of tokens does serve as a measure of total length of the output given by participants. It can be argued that if participants are given the same task, which in this case consisted in retelling as much as they recalled from the film excerpt, participants who give longer answers are performing better than participants that are giving brief answers. Similarly, if a participant uses a higher number of types, one could argue that s/he is using a higher number of different words for the same retelling task, and hence is retrieving more lexical items for the same task. However, this has been disapproved in the literature (see, e.g., Malvern *et al.*, 2004). However, both tokens and types were counted using the functionality included in the CLAN software.

Type-token ratio is calculated as the ratio between the number of types and the number of tokens (Johnson, 1939). Type-token ratio is not stable if we change sample length, limitation that Johnson solved by using samples of the same number of tokens. In this study, I calculated both: a *sample-size dependent type-token ratio* (TTR), which I calculated dividing the number of types uttered by the number of tokens used; and a *sample-size independent type-token ratio* (TTR2), which I calculated using only the first 300 words uttered by each speaker. Both TTR and TTR2 were calculated with the CLAN software.

The *moving average type-token ratio* (MATTR) is a measure of lexical diversity that solves the TTR sample-size dependency problem by computing TTR through a moving window that is independent of text length (Covington & McFall, 2010). The algorithm gives a value that is usually in the range between 0.6 and 0.8, depending on the window size chosen for a particular analysis. This measure can detect changes within a text, as well as differences between texts. In the present study I used the MATTR calculation program included in the CLAN software, using a window size of 10.

The measure of lexical diversity *vocd*-D is the output given by the program *vocd*, included in the CLAN software (MacWhinney, 2000a). This program computes an algorithm that calculates a series of TTR samplings and curve fittings. Although it is widely used in lexical statistic studies, its reliability and accuracy have been questioned in the literature (e.g., MacCarthy & Jarvis, 2007). This measure was the chosen measure for lexical diversity in Keijzer's study (Keijzer, 2007: 192-195), and it was computed again in the present study.

2.4.2 Zipf's law

For the study of Zipf's law, I followed the methodology in van Egmond and colleagues (van Egmond *et al.*, 2015). One of the research questions addressed in the present study is whether Zipf's law, and particularly the exponents in the law, can be used to predict group membership between attriters, controls and acquirers. But one of the main difficulties when using Zipf's law is that both the fit and the exponent parameters are highly dependent on text size (unlike other measures of lexical diversity, e.g., MATTR). That is the reason why, in order to avoid such sample-size effects, I sampled the texts of each participant, in a way that only the first x words of every participants were analyzed.

	300 tokens	400 tokens	500 tokens
attriters	42 (35.9%)	37 (38.5%)	29 (40.8%)
controls	43 (36.8%)	39 (40.6%)	33 (46.5%)
acquirers	32 (27.3%)	20 (20.8%)	9 (12.7%)
total	117 (100%)	96 (100%)	71 (100%)

Table 2.5: Distribution of participants across group in the different Zipf's law analyses

In a first round, I used 300-token texts, covering the same participants as in the analyses reported so far. This was made in order to maximize the number of participants over the number of tokens. A second round of 400-token texts was analyzed, to see if any difference emerged after changing text size. Lastly, a third round of 500-token texts was analyzed, to check for any tendency resulting from increasing text size, both in the exponent and in the fit to the law. The distribution of participants across these different analyses can be seen in Table 2.5. The motivation for using these three analyses is that Zipf's law's parameters are highly dependent on text size, and it could be that the differences between groups were only observable when using higher sample sizes.

All texts were automatically processed to plot word rank and frequency. These were logarithmically transformed, and a first approximation to parameter calculation was made with linear regression, following Izsak (2006). Zipf's law's fit and coefficient parameters were then estimated with maximum likelihood (see Murphy, 2015 for R package used), using the Zipf-Mandelbrot's formulation (Equation 1.3 in Section 1.3.1). In the following sections I present the results from these estimations and the posterior statistical analyses.

Chapter 3

Results

3.1 Lexical Diversity

3.1.1 Tokens and types

In Table 3.1 we can see that the control group's utterances are on average longer (higher number of tokens) than attriters' responses, which are on average longer than acquirers'. The same can be said about the number of types used.

group	tokens	SD	types	SD
attriters (n = 42)	698.3	288.7	224.7	63.4
controls (n = 43)	742.8	298.0	256.8	70.0
acquirers (n = 32)	448.6	117.8	157.2	29.0
total (n = 117)	646.3	283.8	218.1	70.8

Table 3.1: Average number of types and tokens and standard deviation in the film retelling task among groups

A Kruskal-Wallis H test showed that there was a statistically significant difference in number of tokens in the different groups, $\chi^2(2) = 27.563, p = .000$, with a mean number of tokens of 698.3 for the attriter group, 742.8 for the control group and 448.6 for the acquirer group. Post hoc Dun-Bonferri analysis revealed significant differences between the attriter and acquirer group and the acquirer and control group (both with a $p = .000$). This effect can be seen in Figure 3.1a. Another Kruskal-Wallis H test showed that the differences in number of tokens between genders was not significant ($\chi^2(1) = 2.676, p = .102$). A third Kruskal-Wallis H test showed that there was a statistically significant difference in the number of tokens depending on the level of education, $\chi^2(4) = 25.598, p = .000$. Post hoc Dun-Bonferri analysis revealed a significant difference ($p = .000$) between secondary education basic (mean 44.98) and college (mean 90.77) and a significant difference ($p = .029$) between secondary education plus (mean 59.32) and college (mean 90.97). This can be seen in Figure 3.1c. One last Kruskal-Wallis H test showed that there were significant differences in the number of tokens across regions of origin of participants, $\chi^2(3) = 7.357, p = .025$, but post hoc analysis showed no significant differences among any of the origins. Why this significance was not found can be seen in Figure 3.1e.

A one-way ANOVA was conducted to compare the mean number of types across groups, and it yielded significant results with a Welch's $F(2, 72.06) = 45.19$, $p = .000$. Post-hoc Games-Howell analysis revealed that the same effect was found in the token analysis: an effect between attriters and acquirers and controls versus acquirers, both with a $p = .000$. A close to significant effect was found this time between the attriter and control groups, with a $p = .074$. In Figures 3.1a and 3.1b the distribution of the number of types and tokens of each experimental groups can be observed. After that, a Mann Whitney test did not found significant differences of number of types between masculine and feminine participants. A Kruskal-Wallis H test showed that there was a statistically significant difference in the number of types depending on the level of education, $\chi^2(4) = 30.63, p = .000$. Post hoc Dun-Bonferri analysis revealed a significant difference ($p = .000$) between secondary education basic (mean 43.55) and college (mean 92.06), a significant difference ($p = .006$) between secondary education basic (mean 43.55) and university (mean 76.78), and a significant difference ($p = .018$) between secondary education plus (mean 58.80) and college (mean 92.06). This can be observed in Figure 3.1d. One last Kruskal-Wallis H test showed that there were significant differences in the number of types across regions of origin of participants, $\chi^2(3) = 11.59, p = .009$, and post hoc Dun-Bonferri analysis only showed significant differences among participants from the central part (mean 65.64) and participants from the southern area (mean 78.29). This can be observed in Figure 3.1f.

These results show that merely counting tokens or types is not enough to distinguish the attriter and the acquirer group, even when in the present study it seems useful to distinguish both groups from the acquirer group. Since I was working with texts with different lengths, the main idea that can be gathered from these results is that the texts from the acquirer group are significantly shorter than the texts from both the attriter and control groups, as it can be observed in Figure 3.1a. From these results, no conclusions about the lexical diversity of the text samples can be drawn, but we can rather draw conclusions about the length of the interventions.

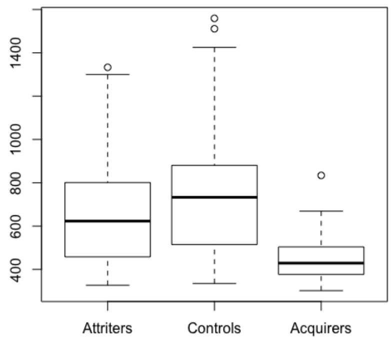
3.1.2 Type-token ratio

The average values of *sample-size dependent type-token ratio* (TTR) and *sample-size independent type-token ratio* (TTR2) across groups can be seen in Table 3.2.

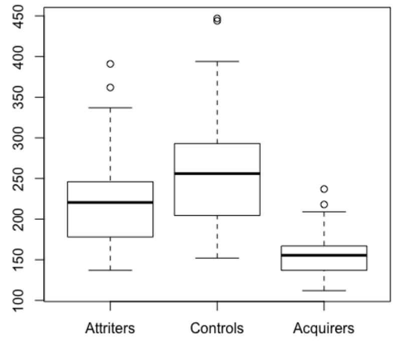
group	TTR	SD	TTR2	SD
attriters (n = 42)	0.340	0.054	0.432	0.035
controls (n = 43)	0.365	0.060	0.460	0.035
acquirers (n = 32)	0.357	0.041	0.394	0.033
total (n = 117)	0.354	0.054	0.432	0.043

Table 3.2: Average values of the sample-size dependent TTR (TTR) and the sample-size independent TTR (TTR2) and standard deviations (SD) in the film retelling task among groups

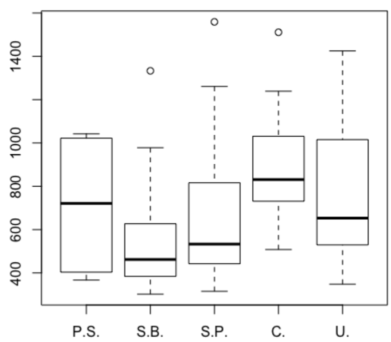
A one-way analysis of variance was carried out to compare the mean values of TTR with the independent variable group. There were no statistically significant differences between group means as determined by one-way ANOVA ($F(2, 114) =$



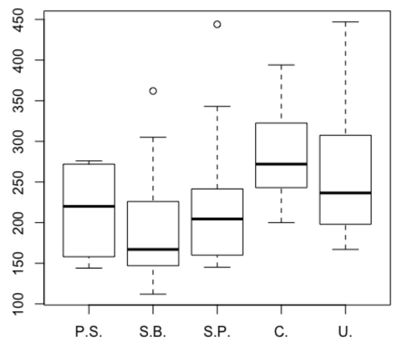
(a) Distribution of tokens across groups



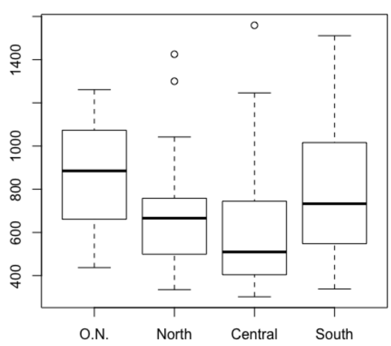
(b) Distribution of types across groups



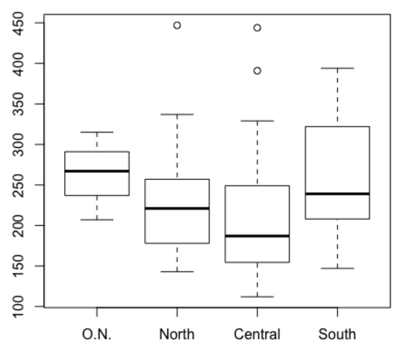
(c) Distribution of tokens across education level



(d) Distribution of types across educational level



(e) Distribution of tokens across area of origin



(f) Distribution of types across area of origin

Figure 3.1: Distributions of tokens and types among groups, educational level and area of origin. In row 2, in order: primary school, secondary education basic, secondary education plus, college and university. In row 3, O.N. = participants born outside the Netherlands.

$2.467, p = .089$). Another one-way analysis of variance was carried out to compare the mean values of TTR2 with the independent variable group, and this time there was a statistically significant difference between groups as determined by an $F(2, 114) = 32.761, p = .000$. The Games-Howell post-hoc procedure revealed that this effect existed for the attriter group versus the control group ($p = .002$), for the attriter group versus the acquirer group ($p = .000$) and for the control group versus the acquirer group ($p = .000$). This difference of significance between TTR and TTR2 can be observed in Figures 3.2a and 3.2b.

No sociolinguistic variable was found to be significant in relation to the sample-size dependent TTR scores in the task. TTR across educational level and place of origin can be observed in Figures 3.2c and 3.2e, respectively. No significant results were found for the gender variable for sample-size independent type-token ratio (TTR2). However, there was a small effect of educational background on TTR2, with an $F(4, 14) = 2.625, p = .038$, being the difference significant. Games-Howell post hoc analysis revealed an almost significant difference between secondary basic education and university ($p = .066$) and between second basic education and college ($p = .073$). This can be seen in Figure 3.2d. The tendency is the same tendency that was attested in earlier analyses, and also in Keizer (2007: 194), and leads to the following conclusion: the higher someone's educational background is, the higher are their diversity scores. This tendency can be observed in mean values in Table 3.3, with the exception of primary school participants, with an average TTR2 of 0.428; which is higher than secondary school basic participants, with an average TTR2 of 0.421. Such irregularity is probably due to the small number of participants with only primary school, which is particularly small (n = 4) when compared to the following groups (n = 53, n = 28, n = 16...).

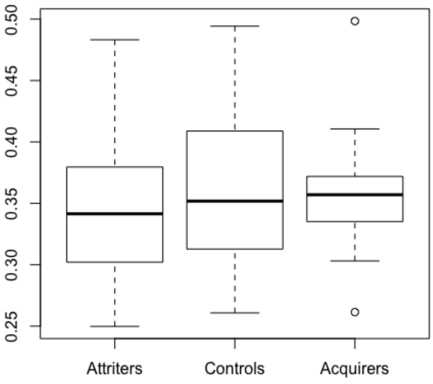
	TTR2	SD
primary school (n = 4)	0.428	0.032
secondary school basic (n = 53)	0.421	0.041
secondary school plus (n = 28)	0.429	0.049
college (n = 16)	0.451	0.038
university (n = 16)	0.452	0.037
total (n = 117)	0.432	0.043

Table 3.3: Average values and standard deviation of the sample-size independent TTR (TTR2) and standard deviations (SD) across different education levels

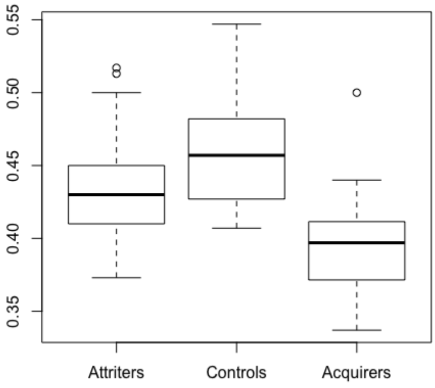
A one-way analysis of variance showed that there was a significant difference in TTR2 group means according to the origin of participants, with an $F(3, 113) = 3.90, p = .011$. A Games-Howell post hoc analysis showed that this difference was significant ($p = .012$) between participants from the central part (mean 0.42) and participants from the southern part (mean 0.56). These differences can be observed in Figure 3.2f.

3.1.3 Moving average type-token ratio

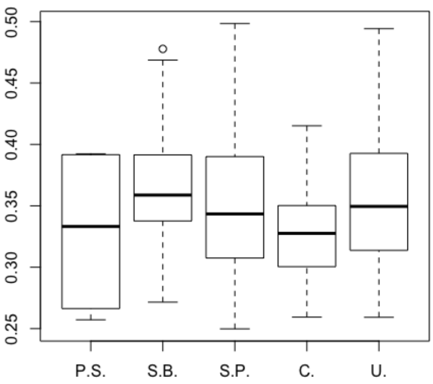
The overall results for MATTR can be seen in Table 3.4. An analysis of variance calculated on the means of the different groups showed that the means were significantly different, with a Welch's $F(2, 67.27) = 18.03, p = .000$. A post hoc



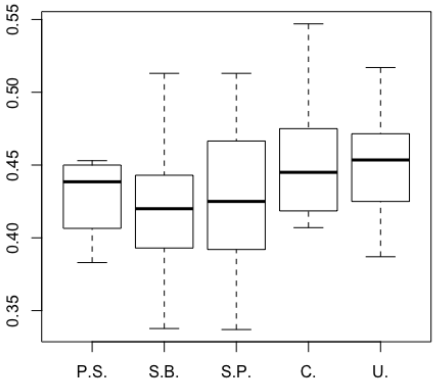
(a) TTR across groups



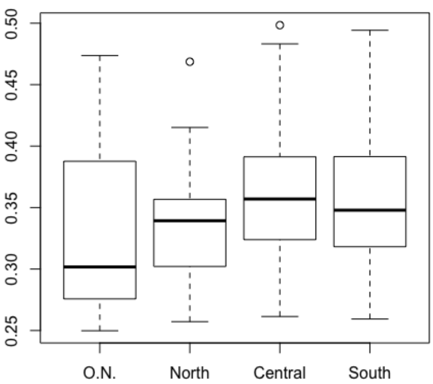
(b) TTR2 across groups



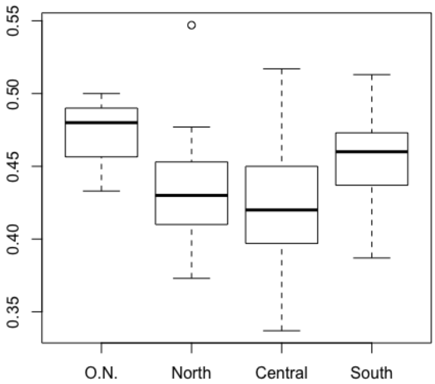
(c) TTR across education level



(d) TTR2 across educational level



(e) TTR across area of origin



(f) TTR2 across area of origin

Figure 3.2: Distributions of *sample-size dependent type-token ratio* (TTR) and *sample-size independent type-token ratio* (TTR2) among groups, educational level and area of origin. For abbreviations see Figure 3.1

Games-Howell analysis showed that the difference was significant between the attriter and the control group, between the control group an the acquirer group, and between the attriter group and the acquirer group, in all three cases with a $p = .000$ (see Figure 3.3a).

	MATTR	SD
attriters (n = 42)	0.903	0.025
controls (n = 43)	0.928	0.014
acquirers (n = 32)	0.908	0.020
total (n = 117)	0.914	0.23

Table 3.4: Average values of the moving-average type-token ratio (MATTR) among groups

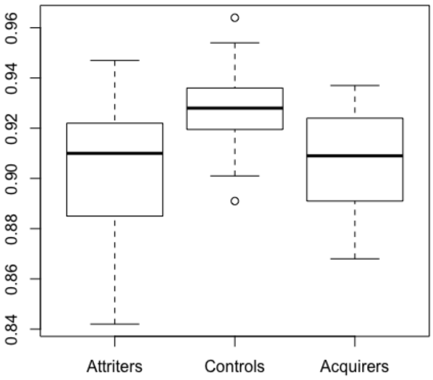
Another ANOVA was calculated to compare mean differences between gender groups and the differences were found significant with a Welch's $F(1, 105.91) = 5.66, p = .019$, showing that the difference between males (mean 0.909) and females (0.911) was statistically significant. Two more analyses of variance were done for the variables education and area, but they did not yield significant results (see Figures 3.3c and 3.3e, respectively).

3.1.4 *vocd*-D

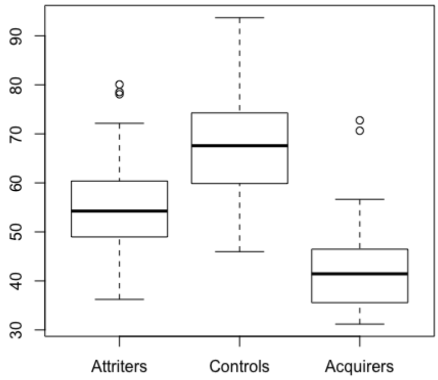
The overall results for *vocd*-D can be seen in Table 3.5. A Kruskal-Wallis H test showed that there was a statistically significant difference in *vocd*-D in the different groups, $\chi^2(2) = 58.90, p = .000$. Post hoc Dun-Bonferri analysis revealed significant differences between the attriter and acquirer group, the attriter and control group, and the acquirer and control group (all three with a $p = .000$, cf. Table 3.5). This effect can be seen in Figure 3.3b. An analysis of variance showed that the difference between genders was not significant. Another Kruskal-Wallis H test showed that there was a statistically significant difference in *vocd*-D depending on the level of education, $\chi^2(4) = 15.09, p = .005$. Post hoc Dun-Bonferri analysis revealed a significant difference ($p = .014$) between secondary education basic (mean 52.5) and college (mean 64.9) groups. This can be seen in Figure 3.3d. One last Kruskal-Wallis H test showed that there were significant differences in *vocd*-D across regions of origin of participants, $\chi^2(3) = 13.67, p = .003$, and post hoc Dun-Bonferri analysis showed that this difference was significant ($p = .002$) between the central part (mean 54.0) and the southern part (mean 66.9) and almost significant ($p = .060$) between the northern part (mean 56.1) and the southern part (mean 66.9). These effects can be seen in Figure 3.3f.

	vocd-D	SD
attriters (n = 42)	55.6	10.3
controls (n = 43)	67.3	10.5
acquirers (n = 32)	43.1	9.5
total (n = 117)	56.5	13.3

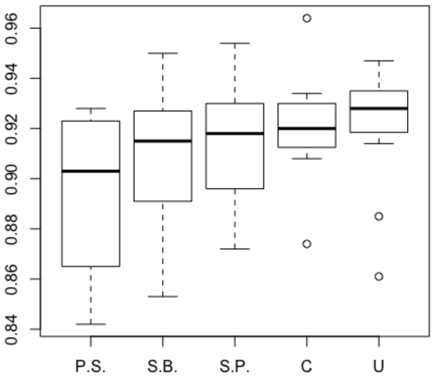
Table 3.5: Average values of *vocd*-D among groups



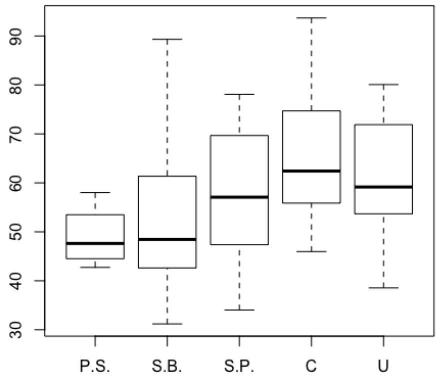
(a) MATTR across groups



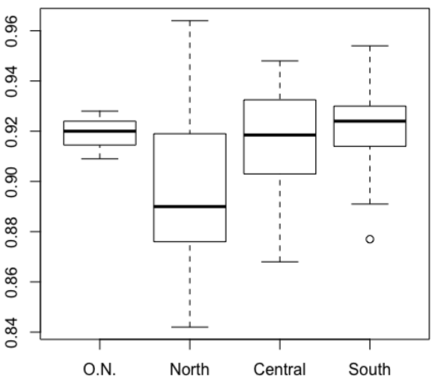
(b) *vocd*-D across groups



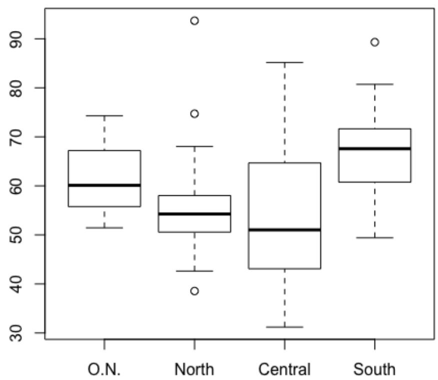
(c) MATTR across education level



(d) *vocd*-D across educational level



(e) MATTR across area of origin



(f) *vocd*-D across area of origin

Figure 3.3: Distributions of MATTR (left) and *vocd*-D (right) among groups, educational level and area of origin. For abbreviations see Figure 3.1

3.1.5 Summary of lexical diversity results

In this section I applied different measures of lexical diversity to the same spoken corpus. In the original study, Keizer compared *vocd*-D to the sociolinguistic variables of group, gender, educational level, region of origin and age, although this last one was excluded from the present study. Keijzer found that *vocd*-D had predictive power for experimental group (i.e., attriters, controls or acquirers) membership (Keijzer, 2007: 192-195). In this study, only three measures had the same predictive power (i.e., TTR2, MATTR and *vocd*-D), and number of tokens, number of types and TTR were not able to predict group membership. Keijzer also found that *vocd*-D was able to predict differences between participants from the north and participants from the south as well as between participants from the central part and participants from the south. Besides *vocd*-D, no measure used in this study was able to significantly predict the same. The newly calculated *vocd*-D predicted the same significant differences, plus a difference between north and center, which, however, was only close to reaching statistical significance. On the other hand, TTR2 did not predict very different results from those of *vocd*-D, which suggests that keeping sample size constant might be enough to overcome type-token ratio's sample-size dependency. Surprisingly, MATTR did not have any predictive power for area of origin differences, which suggests that *vocd*-D might be a better choice. Anecdotally, it did predict differences in gender that no other measure predicted.

3.2 Zipf's Law

3.2.1 Fit

Average, minimum and maximum values of the fitting coefficients are given in Table 3.6. Generally, a good fit to Zipf's law is observed, with an $R^2 \geq 0.906$ for all cases. This value corresponds to the minimum value of the acquirer group in the 300-token analysis. On the other hand, the highest value is the maximum value found for the control group in the 500-token analysis, with a $R^2 = 0.993$. In no analysis was this fit higher for the control group than for the attriter group, but the R^2 values are on average higher for attriters than for acquirers. I tested whether these differences were significantly different by means of Kruskal-Wallis H tests, and these differences became more significant as the number of tokens was increased. In the 300-token analysis, the differences between attriters (mean 0.97), controls (mean 0.96) and acquirers (mean 0.97) were not significant, with a $\chi^2(2) = 1.77, p = .414$. In the 400-token analysis, these differences become almost significant, with a $\chi^2(2) = 5.93, p = .052$, between attriters (mean 0.98) and controls (mean 0.97). One last Kruskal-Wallis H test showed that there were significant differences in R^2 across groups in the 500-token analysis, with a $\chi^2(2) = 7.34, p = .026$, and post hoc Dun-Bonferri analysis showed that this difference was significant ($p = .002$) between the attriters (mean 0.98) and controls (mean 0.97), but not significant between these two groups and the acquirer group (mean 0.98). The differences across groups in the 300-token, 400-token and 500-token analyses can be seen in Figures 3.4a, 3.4c, and 3.4e, respectively.

R^2	300 tokens			400 tokens			500 tokens		
	attriters	controls	acquirers	attriters	controls	acquirers	attriters	controls	acquirers
mean	0.967	0.962	0.966	0.977	0.969	0.972	0.981	0.974	0.976
min	0.991	0.913	0.906	0.944	0.926	0.918	0.954	0.942	0.954
max	0.985	0.989	0.987	0.993	0.990	0.989	0.992	0.993	0.990
SD	0.015	0.018	0.019	0.011	0.014	0.170	0.008	0.012	0.011

Table 3.6: Average, minimum and maximum values of R^2 and standard deviation across groups in the different fits of the different experiments

α	300 tokens			400 tokens			500 tokens		
	attriters	controls	acquirers	attriters	controls	acquirers	attriters	controls	acquirers
mean	0.851	0.822	0.989	0.895	0.855	0.939	0.930	0.890	0.986
min	0.692	0.711	0.724	0.773	0.743	0.817	0.816	0.811	0.877
max	1.017	1.042	1.025	1.119	1.014	1.019	1.110	1.020	1.102
SD	0.079	0.075	0.083	0.075	0.067	0.054	0.074	0.055	0.067

Table 3.7: Average, minimum and maximum values of α and standard deviation across groups in the different fits of the different experiments

3.2.2 Slope

Average, minimum and maximum values of the α exponent are given in Table 3.7. In all three analyses we see that the same relation holds: values of α are higher for acquirers, followed by attriters and then controls. These differences were statistically tested to find if they were significant. In the 300-token analysis, a one-way ANOVA showed that the differences in α across groups were significant, with an $F(2, 114) = 8.56, p = .000$. Post-hoc Games-Howell analysis revealed that these differences were significant ($p = .040$) between attriters (mean 0.85) and acquirers (mean 0.90), they were also significant ($p = .000$) between controls (mean 0.82) and acquirers, but not significant between attriters and controls ($p = .213$). A second one-way ANOVA was tested to study the differences in average values of α between groups in the 400-token analysis, and significant differences were found, with an $F(2, 93) = 10.53, p = .000$. Post-hoc Games-Howell analysis revealed that there were significant differences between all groups. First, there was a significant difference ($p = .046$) between attriters (mean 0.90) and controls (mean 0.86). The differences between controls and acquirers (mean 0.94) was also signficant ($p = .000$), and so was the difference between attriters and acquirers ($p = .033$). One last one-way ANOVA was calculated to study the differences in average values of α between groups in the 500-token analysis, and significant differences were found, with an $F(2, 68) = 8.39, p = .001$. Post-hoc Games-Howell analysis revealed that there were significant differences ($p = .001$) between controls (mean 0.89) and acquirers (mean 0.99), almost significant differences ($p = .058$) between attriters (mean 0.93) and controls, and close to significant differences ($p = .081$) between attriters and acquirers. The differences between groups in the 300-token, 400-token and 500-token analyses can be seen in Figures 3.4b, 3.4d, and 3.4f, respectively.

As discussed before, the expected value for the exponents of the law in Formulation 1.3 (cf. Section 1.3.1) are $\alpha \approx 1$ and $\beta \approx 2.7$ (Mandelbrot, 1953, 1961; Zipf, 1936, 1949). In the results of the experiment I only found that α reached those values in the maximum values of each experiment. Mean values of control speakers do not come close to 1, and this is due to the use of small sample sizes. As we can see in Table 3.7, this value tends to increase when the number of tokens increases. For example, for controls, the mean value is 0.82 in the 300-token analysis, 0.86 in the 400-token analysis and 0.89 in the 500-token analysis. Using these small sizes makes these values difficult to interpret by themselves, but what remains important is that the differences between groups within same token-size analyses is statistically different.

Average, minimum and maximum values of the β exponent are given in Table 3.8. As a reminder, these β values do not correspond to the β formulation (Formulation 1.6 in Section 1.3.1), so they cannot be compared to α, but they correspond to the Zipf-Mandelbrot formulation (Formulation 1.3 in Section 1.3.1). As we can see in the table, they strongly differ from the expected $\beta \approx 2.7$. The β exponent is mainly based on the first few ranks of the distribution, and these few ranks are calculated on texts with very sample sizes for β to be statistically reliable. In Table 3.8 we can see that there is an observable tendency in the value, i.e., it tends to increase when the number of tokens used is higher, becoming closer to the ideal $\beta \approx 2.7$. Between groups, we can see that in the 300-token and 400-token analyses the value is smaller in the acquirer group, followed by the attriter group and followed by the control group, in a similar pattern to α. However, the exact opposite tendency is found in the 500-token analysis. Despite being statistically unreliable, I tested the differences between groups in all three analyses (one way ANOVA for the 300-token analysis and Kruskal-Wallis H tests for the 400-token and the 500-token analyses) and in no case were these differences statistically significant.

β	attriters		controls		acquirers	
	μ	SD	μ	SD	μ	SD
300 tokens	1.179	0.982	1.372	1.004	0.998	0.825
400 tokens	1.452	1.555	1.486	0.904	1.170	0.711
500 tokens	1.647	0.971	1.644	0.686	1.847	0.657

Table 3.8: Average values of the exponent β and distribution across groups in 300-token, 400-token and 500-token analyses.

3.2.3 Zipf's law and the sociolinguistic variables

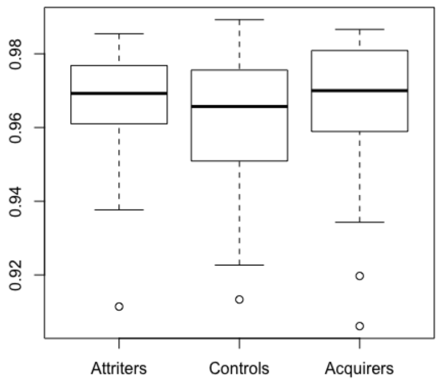
Besides group membership, I tested whether there was an effect of other sociolinguisic variables (namely gender, education level or area of origin) in the results of R^2, α, and β, but there was no significant interaction, regardless of the number of tokens used in the analysis, i.e., there was no significant interaction in neither the 300-token, the 400-token, nor the 500-token analysis.

3.3 Conclusions

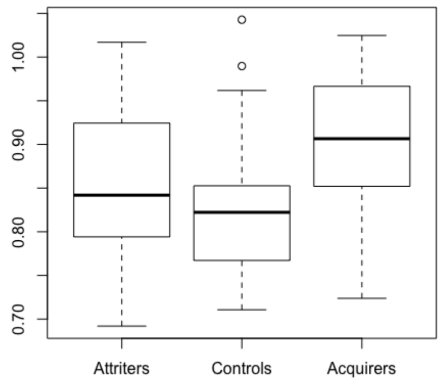
In the Theoretical Introduction, we saw that attriter speech usually consists in the use of a smaller amount of lexical items, with frequency distributions that usually consist in highly frequent words being used more often, and low frequency items being less common. In this study I aimed to analyze this phenomenon by comparing traditional measures of lexical diversity and the use of Zipf's law. Particularly, I aimed to give answer to the following research questions:

1. Does attriter speech conform to Zipf's law?
2. If (1) holds, is it with the same or a different slope to controls (and acquirers)?
3. Is the exponent in Zipf's law a strong predictor for group membership when compared to other measures of lexical diversity?

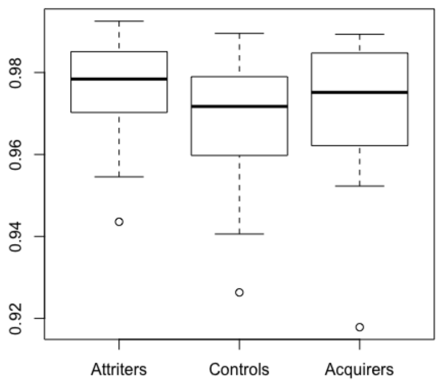
The answer to Question (1) is given by the results in the fit coefficient R^2 that was calculated with linear regression and maximum likelihood estimation in Section 2.2.1.1. We saw that generally there was a good fit to the law given by this value in all different analyses and for all groups, with an overall $R^2 \geq 0.906$ for all cases. We also saw that in no analysis was this fit higher for the control group than for the attriter group, but the R^2 values are on average higher for attriters than for acquirers in all analyses, and significantly higher for the attriter group than for the control group in the 500-token analysis. So we can confirm that attriter speech does conform to Zipf's law, and whether it does so better than controls will be discused in the next chapter.



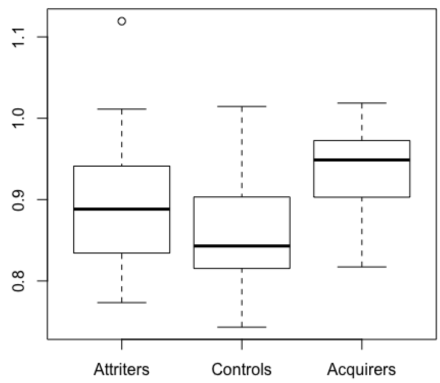
(a) R^2 in 300-token analysis



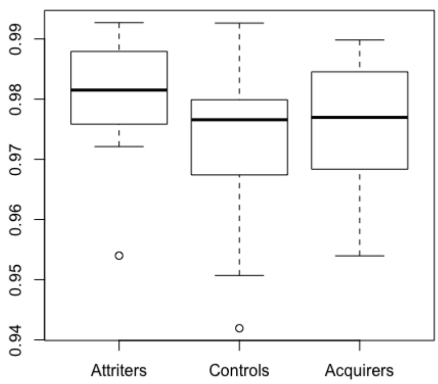
(b) α in 300-token analysis



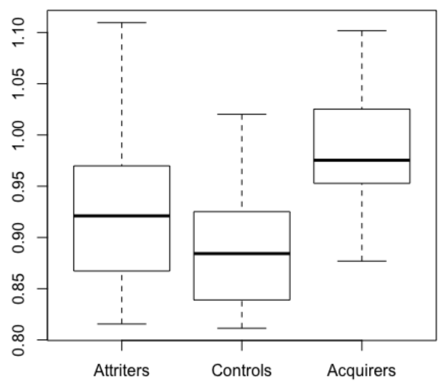
(c) R^2 in 400-token analysis



(d) α in 400-token analysis



(e) R^2 in 500-token analysis



(f) α in 500-token analysis

Figure 3.4: Distributions of R^2 (left) and the slope α (right) across groups in the 300-token, 400-token and 500-token analyses.

Like it was explained before, the β exponent in Zipf-Mandelbrot's formulation is mainly based on the first few ranks of the Zipfian distribution, and these few ranks are too small in small sample sizes to be statistically reliable. This is why Questions (2) and (3) will be answered only using exponent α. Regarding Question (2), I found that there was a difference in slope between groups, with α being the highest in the acquirer group, followed by the attriter group and followed by the control group in all analyses. These differences were only statistically significant in the 400-token analysis, almost significant for all groups in the 500-token analysis (and significant between controls and acquirers), and only significant between acquirers and attriters and acquirers and controls in the 300-token analysis. These exponents seem to be highly dependent on text size, and this dependency can be in the obtained values across 300-token, 400-token and 500-token analyses: the exponents become closer to 1 as the number of tokens used is increased. It seems to be the case that acquirers tend to show the highest slope among the three groups, followed by attriters, and followed lastly by controls. In other words, acquirers' word frequency distributions tend to show a steeper slope, followed by attriters', which present a slightly less steep slope, and lastly followed by controls', who present the least steeper slope.

In Figure 3.5, I present the frequency distributions of participants BMT, MIS and NAF, who happen to be the closest representants to the average α values in the attriter, control, and acquirer group, respectively. If we look at the frequency-rank plots, we can see that NAF (last row), the average acquirer, makes a more frequent use of the first ranks than the average attriter (in this case BMT, in the first row), and this can be (barely) observable between the average attriter (BMT) and the average control (MIS, row 2). We can also observe that the average acquirer makes less use of hapax legomena than attriters and controls. The same tendency can be observed, although with some difficulty, in the log rank - log frequency plots: the line that appears in each plot is a linear representation of the slope α. This line is steepest in the acquirer, followed by the attriter and followed by the control.

In order to answer Question (3), we can observe that this steepness coincides with the results obtained from *vocd*-D scores, so in a way α predicts group membership. I found that *vocd*-D scores were higher in controls, followed by attriters and acquirers. The higher the proficiency of the speaker is, the steeper is the slope of their frequency distributions (and consequently the smaller the α value is). This confirms that the slope of Zipf's law reflects the lexical diversity of the sample, which can be explained intuitively: the more lexically diverse a text is, the less common words it contains, wich are usually function words. Similarly, the more lexically diverse the text in question is, the more hapax legomena it will contain, which are usually content words of a highly restrictive semantic reference. This results in a steeper word frequency distribution, and a smaller α value. But is the α exponent a reliable measure to assess lexical diversity? In the present study, it had predictive power to significantly distinguish controls and acquirers in all analyses, it predicted the difference between attriters and acquirers in the 300-token analysis and the 400-token analysis (and almost significantly in the 500-token analysis), and again between attriters and controls in the 400-token analysis and almost significantly in the 500-token analysis, but not in the 300-token analysis. These differences can be explained with the different number of participants used in the different analyses and in the text-size dependency of the α value.

In order to answer whether Zipf's law is a strong predictor for group membership

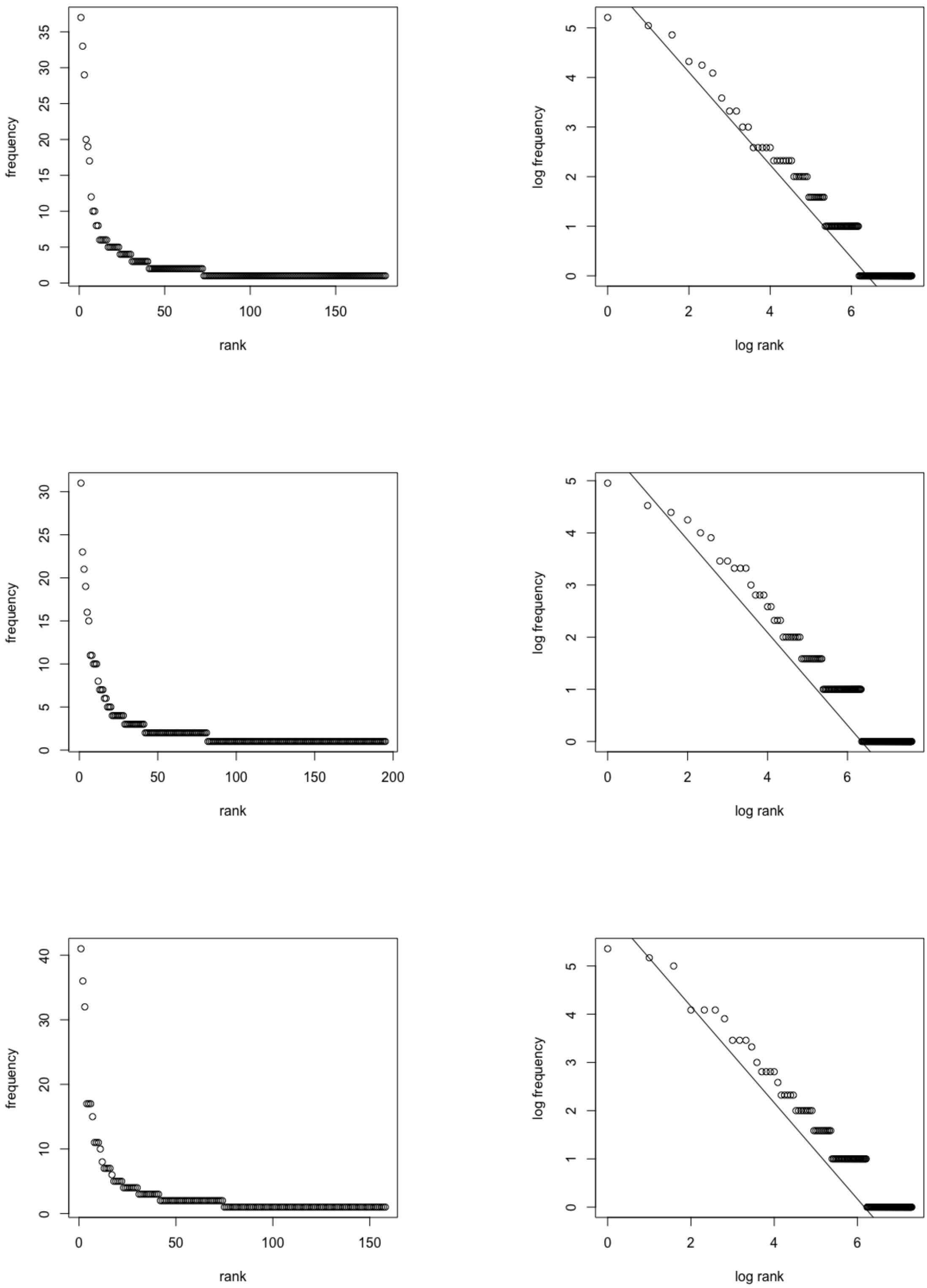


Figure 3.5: Frequency - rank distributions and log frequency - log rank distributions (i.e., α distributions) of participants BMT ($\alpha = 0.93$), MIS ($\alpha = 0.89$) and NAF ($\alpha \approx 1$), from the attriter, control and acquirer group, respectively.

when compared to other measures of lexical diversity (Question 3), it needs to be compared to other measures of lexical diversity. Particularly, we can compare it with the measures that were calculated in the present study: number of types, number of tokens, type-token ratio, text-size independent type token ration (TTR2), moving average type-token ratio (MATTR), and *vocd*-D. In Table 3.9 we can see a relation of all the significant or almost significant results where measures of lexical diversity could predict either participants' group membership, or other sociolinguistic variables like gender, educational level or place of origin. As it can be seen in the table, Zipf's law α exponent is a better predictor than the regular type-token ratio (TTR), but fit has no predictive power at all. Overall, the exponent of Zipf's law has less predictive power than other measures of lexical diversity like *vocd*-D, MATTR, or even text size-independent type-token ratio (TTR2). This may be the result of the exponent's dependency to text-size. If the exponent was to be used in assessing lexical diversity, it would require the use of longer texts. This makes the exponent a much less powerful tool than text-size-independent measures like MATTR, *vocd*-D or even TTR2, even when the reliability of these measures is already questioned in the literature. The answer to the Question 3 is that, even when α predicts group membership to some extent, it does not perform as well as other lexical diversity measures, and consequently it is *not a strong predictor* when compared to these.

	Group			Gender	Education level			Region	
	attr./contr.	attr./acq.	contr./acq.		S.B.-C.	S.B.-U.	S.P.-C.	N-S	C-S
Tokens	-	$p = 0.000$	$p = 0.000$	-	$p = 0.000$	-	$p = 0.029$	-	-
Types	$p = 0.074$	$p = 0.000$	$p = 0.000$	-	$p = 0.000$	$p = 0.006$	$p = 0.018$	-	$p = 0.009$
TTR	-	-	-	-	-	-	-	-	-
TTR2	$p = 0.002$	$p = 0.000$	$p = 0.000$	-	$p = 0.073$	$p = 0.066$	-	-	$p = 0.011$
MATTR	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.019$	-	-	-	-	-
vocd-D	$p = 0.000$	$p = 0.000$	$p = 0.000$	-	$p = 0.014$	-	-	$p = 0.060$	$p = 0.002$
α_{300}	-	$p = 0.040$	$p = 0.000$	-	-	-	-	-	-
R^2_{300}	-	-	-	-	-	-	-	-	-
α_{400}	$p = 0.046$	$p = 0.033$	$p = 0.000$	-	-	-	-	-	-
R^2_{400}	$p = 0.052$	-	-	-	-	-	-	-	-
α_{500}	$p = 0.058$	$p = 0.081$	$p = 0.001$	-	-	-	-	-	-
R^2_{500}	$p = 0.002$	-	-	-	-	-	-	-	-

Table 3.9: Significant or close to significant results found in the different analyses performed with the different measures of lexical diversity and the parameters in Zipf's law. Only variables in which at least one of the analyses was found significant are shown.

Chapter 4

Discussion

In the previous section I concluded that attriter speech conforms to Zipf's law, and it does so with a different slope than both controls and acquirers. I also found that, when compared to other measures of lexical diversity, the exponents in Zipf's law are not a powerful tool to measure lexical diversity, although they have predictive power when it comes to predicting group membership among attriters, acquirers and controls. In this last section, three main points will be discussed. In the first section, I will combine the results from this and previous studies, paying special attention to the quantitative aspects of the different findings. In the second section, I will attempt to combine these findings and interpret them within the different theories regarding the origin of Zipf's law. In the last section, I will try to connect the findings in Zipf's law studies with the different theories that are traditionally used to explain language attrition, and particularly lexical attrition phenomena.

4.1 Current and Previous Findings

In the current study it was found that attriter speech conforms to Zipf's law with a difference in slope: the calculated slope α was the highest in the acquirer group, followed by the attriter group and with the control group being the lowest of the three. We also saw that these differences were significant in the 400-token analysis, but only almost significant in the 500-token analysis. I explained that, while the quantitative differences among groups were intensified as the sample size was increased, not gaining statistical significance in the 500-token analysis can be ascribed to the reduced number of participants in such analysis.

On the other hand, in the experimental results from the present study, we saw that α only reached the ideal value $\alpha \approx 1$ in the maximum values of each experiment. Quantitatively, these values can be compared with the α values reported in aphasic speech and in the development of child language (Baixeries *et al.*, 2013; van Egmond *et al.*, 2015). This difference between the reported and the ideal values can be attributed to two different factors. On one hand, we saw that the parameters of Zipf's law are highly dependent on text size, so smaller-than-normal values of α are expected when the number of tokens included in the present experiment is reduced. Both Baixeries and colleagues and van Egmond and colleagues used small sample sizes: Baixeries and colleagues carried out 250-word and 500-word analyses and found an average $\alpha = 0.75$ for children and an $\alpha = 0.65$ for investigators, who were tested to compare child speech to an adult speech that did not interact with the children in development (Baixeries *et al.*, 2013); whereas van Egmond and colleagues carried out a 352-token analysis and found an average $\alpha = 0.83$ for aphasics and an $\alpha = 0.68$ for control speakers (van Egmond et al, 2015). This tendency was observable in the experimental results: when I increased number of tokens the value of α came closer to the ideal $\alpha \approx 1$ (see Table 3.4). So it seems manifest that the standard assumption of $\alpha \approx 1$ can only be made for lexical analyses of large corpora, since this exponent is extremely reduced for smaller texts. On the other hand, the differences in α can be explained with the use of spoken instead of written data: the use of spoken data might include more discourse linkers or common phrases than written texts, but spoken corpora often consist of smaller text samples than written corpora. These two factors display the importance of using control groups to account for these deviations from the traditional values: analyzing the values per se yields no significant information, but it is the differences between groups that should be compared.

In the present study, we see that the values of α can be correlated to the *vocd*-D scores reported by Keijzer (2007), which were also corroborated by the results in the present analysis. In all analyses, I found that the exponent had the lowest values for the control group, followed by the attriter group and lastly by the acquirer group. Similarly, D scores are the highest for the control group, followed by the attriter group and lastly by the acquirers. I computed a Pearson product-moment correlation coefficient to assess the relationship between the α exponents and the *vocd*-D scores in the 400-token analysis, and there was a negative correlation between the two variables, $r = -0.572$, $n = 117$, $p = 0.000$. This correlation can be observed in the scatterplot in Figure 4.1. One possible explanation for this correlation would be that both *vocd*-D and α were measures of lexical diversity. This conclusion is a priori intuitively correct. Nevertheless, I found that these *vocd*-D scores also had predictive power for other sociolinguistic variables like place of origin and education level (see Section 3.1.4), similar to the findings in Keijzer (2007: 192-196). If both α and *vocd*-D can be used as measures of lexical diversity, why *vocd*-D performs as a measure to identify other sociological variables like place of origin or education level (whereas α does not) still needs to be answered. Particularly, I found that *vocd*-D scores of highly educated participants were higher than those of participants with a lower education, and I also found higher *vocd*-D scores in participants from the south. There is no a priori reason to believe that participants from the south should perform better than other participants in lexical diversity, so these differences could be seen as accidental. However, the differences regarding educational background correspond to the performance of the C-test: highly educated participants perform better in both lexical diversity and overall language proficiency. Educational background is regarded as a determining factor for the outcomes of language attrition (see, e.g., Köpke & Schmid, 2004), and while α predicts differences between attriters, controls and acquirers in the present study, it does not predict educational background. Why the results in α do not capture these differences could be an indicator that α does not merely represent lexical diversity, but is the result of a more intricate property of language production.

Baixeries and colleagues compared obtained exponent values with a measure of syntactic complexity, the mean length of utterances (MLU, Baixeries *et al.*, 2013). MLU serves as an estimator for syntactic complexity, and it is a measure that relies on the fact that shorter sentences tend to be simpler. In their analysis of the

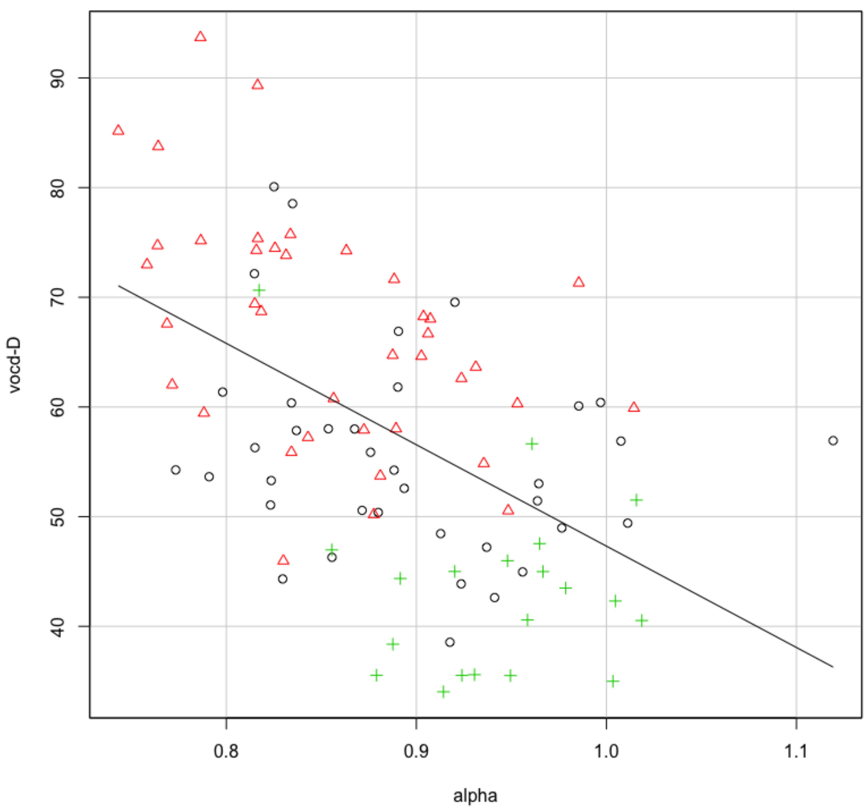


Figure 4.1: Distribution of *vocd*-D scores plotted against the calculated α values in the 400-token analysis, and sorted by groups: attriters (circles), controls (triangles) and acquirers (crosses).

evolution of the exponent in child speech, they found that α tends to decrease with age for many children, and this decrease in the exponent is correlated with an increase in syntactic complexity, measured in MLU. In Section 2.2.2, we saw that, in Keijzer (2007), the participants whose data was used in the present study were tested by means of a C-test to assess their overall proficiency in their L1. She found that there were significant differences among groups, with controls significantly outperforming both attriters and acquirers, and attriters significantly outperforming acquirers (Keijzer, 2007: 184). Taken together, the results from both studies could suggest that the significant differences that were found for α between groups in this study could be related to the syntactic performance of participants, observable in the overall performance on the C-test. However, besides measuring overall language proficiency and lexical diversity, Keijzer also assessed both morphological and morphosyntactic proficiency, measuring specific aspects of language performance, like noun and phrase morphology, negation, passive constructions, and subordination. On one hand, she found that there were significant differences in the morphological and morphosyntactic performance of groups that resembled the overall language proficiency: although the divergences between conditions were not constant, she found that there was a pattern for the scores of both acquirers and attriters, which were typically lower than the scores of controls (id.: 184-255). On the other hand, she did not find significant differences between acquirers and attriters. Using the data from the same participants, I found significant differences in the exponent between attriters and acquirers that were not found in their morphological or morphosyntactic performance. This suggests that the correlation between the exponent and syntactic complexity should be revised, since the correlation between MLU and α could result from lexical diversity being an epiphenomenon of overall language proficiency. If both α and MLU derive directly from the complexity of the language system in terms of lexical diversity and syntactic complexity, respectively, they could both be a measure of the same underlying phenomenon; in this case, language proficiency. Future research should aim to link morphosyntactic complexity and the exponent in Zipf's law by taking actual morphosyntactic phenomena into account, and preferably using natural, free-speech data.

In the study of (linguistic) pathologies, values of the exponent were also reported for schizophrenia and aphasia. All these followed a tendency to increase in α for more repetitive or less lexically diverse speech. In all cases, when α is higher than controls or than the ideal $\alpha \approx 1$, speakers make less use of low-frequency vocabulary in elicited speech. The relevance of the exponent depends on what is assumed to be the origin of Zipfian distributions in human language.

4.2 Revisiting the Origin of Zipf's Law

The potential meaningfulness of the exponent in Zipf's law is tightly related to the different theories that have been postulated to explain the origin of the law. In Section 1.3.2., I pointed out that there was no consensus regarding the origin of the law, but the theoretical frameworks fit into two groups: those which explain the origin of the law via statistical processes, and those which assert that it originates in psychological processes.

The two statistical accounts of the law explained in the Theoretical Introduction were random typing models (also known as intermittent silence accounts) and simple stochastic models. Both these models are problematic in that they account for neither meaningful production of language, nor for the relationship between the meaning of words and their position in relative usage frequency. In other words, they do not explain why function words are more frequent than content words, but they rather predict that frequency of usage is, in a way, proportional to word length. In this study, I found that the three groups conform to Zipf's law, and that there is a significant difference in the slope between groups. These models do not predict these differences, because, if words are generated at random, there would be no reason for acquirers to generate significantly more short words than attriters, who also generate significantly more short words than controls. In this sense, probabilistic accounts lack explanatory power for the current results, and to my knowledge they do not fully elucidate the variation in the exponent in the studies of Zipf's law mentioned before.

Communicative accounts claim that Zipf's law maximizes efficiency and minimizes communication costs, and the law is the result of how lexical items are connected in the lexical network, which results in low-frequency words being usually less syntactically relevant and of a higher semantic content. In this aspect, that they consider the connectivity between linguistic elements to be essential to the derivation of the law makes it an explanatory framework for the meaningfulness of Zipf's law and its exponent. In the models of the law proposed by communicative accounts (e.g., Ferrer i Cancho & Solé, 2003), speaker effort is proportional to the diversity of signals they must use, and listeners pay a cost that is proportional to the entropy over referents given a word. This trade-off of costs between speaker and listener is measured with the parameter λ. This λ has a particular value, $\lambda = 0.41$, in which the optimization model recovers Zipf's law. In this framework, a closer approximation to the law (or a better fit) would mean that this trade-off cost between speaker and listener comes closer to this ideal λ value.

In the present study I found that, in the 400-token analysis and the 500-token analysis, the fit of the law was significantly different between attriters and controls, with R^2 being higher for attriters than for controls in all analyses. This particularity could not be extended to the acquirer group, where there was no pattern, but is similar to the results found in the study of Zipf's law in aphasia, where aphasics showed a higher R^2 than controls in all analyses, with the differences being significant in the token analysis of lexical words (van Egmond *et al.*, 2015). Although it seems difficult to draw conclusions from these results, it could be explained in a communicative framework. It could be that the word-concept associations in aphasic and attriter lexica were altered in a way that their output conformed better to Zipf's law than normally-developed speech. This particularity should be studied further in future research. M. van Egmond (personal communication, June 22, 2016) believes this could also be explained with aphasics' problems with function words: the first ranks of the law are usually function words, whereas the last ranks are mostly restricted-content words. In healthy speech, Zipf's law overestimates the frequency of the first few ranks, and aphasics conforming better to Zipf's law could result from their problems with function words. The fact that it was found that attriter speech conformed better to the law than control speech (and with no significant differences between attriters and acquirers or between acquirers and controls) could be a sign that attriters experience a difficulty with function words similar to aphasics.

However, there are some limitations to the communicative accounts of the law.

First, they assume that all meanings are equally likely to be conveyed (e.g., Ferrer i Cancho & Solé, 2003), which is problematic because this leaves out the topic effects on discourse. Later communication optimization models do not make this assumption, but still fail to incorporate known factors like memory latency (cf. savings effect), frequency effects, and context-based expectations. One could say that a confined topic, as in a film-retelling task, is semantically restricted enough to make such an assumption about its use, but that would still not explain the emergence of the law in the distribution of words for objects with a fixed referential content. Future research should aim for a more holistic account of the law in word frequencies.

We could also compare the current results with the results found in the study of aphasia. Van Egmond and colleagues (van Egmond *et al.*, 2015) hypothesized that the origin of the law is either in the processing of words or in the storage of words. Following the general consensus in the literature, they ascribed aphasic difficulties in word retrieval to problems of processing, rather than storage. They found that aphasic speech conformed to Zipf's law, with a slope (α) higher in aphasics than in controls. This finding suggests that the origin of the law is in the mental lexicon, rather than in the selection and retrieval mechanisms. If we were to assume that attriters' word finding difficulties are in the processing mechanisms (due to increased activation thresholds, for instance), then the results would confirm that, in attriter speech, the law also originates in the mental lexicon. We cannot extend this assumption to acquirer speech, since incomplete acquisition cannot be compared to increased inhibition or breakdown.

My personal intuition about the finding of Piantadosi (2014) that, when learning novel words, participants output them conforming to Zipf's law is that the law might be at work in the normalization of semantic input. This normalization would result in the organization of memory and lexical networks, and could easily explain the prevalence of the law in other human systems like music or computer software. But, in the end, there seems to be no fully satisfactory explanation for the origin of the law, since each theory accounts only partially for the results obtained either in the present or previous studies. It seems relevant to mention a reflection by Piantadosi (2014), in whose view the lack of unitary explanation is probably due to the fact that there are multiple causal forces at play. In this sense, statistical forces and communicative or lexical forces might all operate concurrently in linguistic processes.

4.3 Zipf's Law and Lexical Attrition

In the Theoretical Introduction, we saw that attrition is a phenomenon ascribed to bi- or multilingual speakers, who usually conform to a language system that is different from that of monolingual speakers. The interaction of these two systems is a matter of debate. At the lexical level, theories like decay claim that a dominant L2 may lead to the dissipation of lexical items in L1 because of lack of use of the native language. A different view would assume that these two language systems are active at the same time, and compete during lexical access, leading to bilingual processing requiring a higher cognitive effort. Competing lexical items will be more easily accessed when they are used more frequently, earlier and when they are not inhibited. This results in the reduction of fluency and a slowdown in bilingual lexical access, but is also observable in interferences on the lexical, phonological, and grammatical levels in both languages.

The differences in slope between attriters, acquirers and controls is in line with the characteristics of the speech of each group: a higher value of α indicates that the speakers use fewer different types, and use those types more frequently. These results are in line with early characterizations of attriter speech as evidence of decay, like that of Andersen (1982), who characterized attriter speech as consisting of a reduced number of representations with a smaller variety of lexical items available to them, and those items retained consisting of more common and highly frequent items (Andersen, 1982). However, assuming that the number of representations is reduced would not explain the savings effect: if lexical representations were completely lost, there would be no residual knowledge of such activation, and findings from the savings paradigm would need to be explained (e.g., de Bot et al 2004; MacLeod, 1976). If attriters' lexica was in a way "impaired", the resulting speech would not conform to Zipf's law, assuming that the law originates in the storage of words. In the results it was found that attriters conform to the law significantly better than controls. It could be argued that this could be the result of a bigger (bilingual) lexicon showing more interconnectedness than a monolingual lexicon, and then resulting in a better fit of the law. However, this would not explain why aphasics show a better fit to the law than controls in the study from van Egmond and colleagues (van Egmond *et al.*, 2015). It is very likely that what is impaired in attriter speakers is the access to the phonological form of such lexical forms, in the way that theories like the Activation Threshold Hypothesis claim. The difference in fit can be ascribed to the behavior of the law with the first ranks, as was explained before.

The Activation Threshold Hypothesis and the communicative accounts of Zipf's law could be combined to explain the different findings of the exponent in aphasic, attriter and schizophrenic speech with the communicative notion of connectedness. As mentioned in the Theoretical Introduction, there are different explanations for the lexical access difficulties in anomic aphasia, but most of them rely on the idea of either a reduction in the base level of activation of representations or a delay in the activation (Ewijk, 2013). These notions are similar to the activation slowdown theories of lexical activation in attriter speech and the Activation Threshold Hypothesis itself. The present experimental design does not allow us to make any claim about the first, but we could theorize that the level of activation is correlated to the exponent in which the speech conforms to Zipf's law. This could be explained with the notion of neural connectedness (Ferrer i Cancho, 2006): a higher connectedness would result in an increased possibility of starting from a signal (or stimulus) and reaching the remaining signals and stimuli of the network, which would result in change in the exponent (a smaller α). The result would also be an increased synaptic density. On the other hand, this view would suggest that in attriter speech, where α is higher than controls, the connectedness between the lexical items of the L1 would be reduced, resulting in a decreased synaptic density. This reduction could be explained with L1 items having a higher activation threshold, or with the neural inhibition costs of the L2 elements. Further research on Zipf's law should look at both attriters' L1 and L2, and check if such reduction is still present when speaking the (dominant) L2. This explanation could not be extended to the differences in the exponent found in aphasic speakers: it is generally believed that aphasics have an intact lexicon, but reduced processing resources (e.g., Avrutin, 2006). An intact lexicon would mean that the interconnectedness between the lexical elements is not affected, so the differences in slope found in aphasic speech could not be explained with neural connectedness.

The experimental results in this study can say little about which theoretical framework has a better explanatory power for lexical attrition phenomena. Overall findings in lexical performance of attriters and acquirers does not differ from those reported in Keijzer (2007). She points out that, even when mirror symmetries can be found in the development of attriters and acquirers, the regression hypothesis cannot account for the variation that exists due to external and internal factors in language development, and she opts for usage-based accounts or Dynamic Systems Theory as frameworks in which the regression hypothesis should be redefined. In the Theoretical Introduction, I described the Dynamic Model of Multilingualism (Herdina & Jessner, 2002), in which attrition is seen as the result of not spending enough time or effort in the retention of a language, and being a phenomenon that is directly related to the competition with (an)other language(s). This competition is seen as the responsible factor for the self-organization of resources from unstable to stable (attractor) states. This competition manifests itself as an interplay between system-internal (like intelligence, linguistic capacity, time or motivation) and system-external resources (like language and social input).

Experimental support for the existence of such attractor states is found in Meara (2004), who modeled vocabulary loss in language attrition. As explained in the Theoretical Introduction, he designed lexical networks in which words were connected to two other words. Each element could be in either an activated or in a deactivated mode. In his network, the activation of a particular word depends on the level of activation of the two words it is connected to, so the activation or deactivation of a word would lead to a cascade of activation/deactivation until the network settles in an attractor state. A reduction in the activation of certain nodes or connections in the attriter's lexicon could result in the activation and deactivation of other elements, until the L1 vocabulary reached an attractor state. In this case, if we assumed Steyvers and Tenenbaum's (2005) interpretation of their findings, namely that the organization of the semantic structures is what determines speech to derive in a power law, the exponent of the law could be a result of the change of the connections in the network. In a way, a particular value of the exponent could be the result of an attractor state of a dynamic system.

The same could be argued for the different stages in language acquisition, the self-organizing characteristics of the system can be seen as the cause for the increase in vocabulary size and the decrease in the exponent. As they develop over time, the systems stop in attractor states, which are considered to be preferred, but still unpredictable. This predicts that at different stages in language development, some changes in the exponent could be expected, but the exact variation is unpredictable. The advantage of using a dynamic systems paradigm is that the non-linearity of development can be covered under the same paradigm. In this sense, future research should aim to delineate a holistic dynamic systems model able to account for every possible outcome of language development, including phenomena like acquisition, L2 learning, attrition, and language breakdown. The role of Zipf's law in such a model is still unclear, but it might be connected to the self-organizing principles of vocabulary learning.

Acknowledgements

This research would not have been possible without the help of Merel Keijzer, who made her own research available for the present study. I would like to express my gratitude to her, and also to everyone who participated in her research. I would also like to thank my supervisors Sergey Avrutin and Marjolein van Egmond for their time and dedication. A big thank you to Marjolein for her constant help, useful comments and remarks, and an overall fruitful engagement throughout the whole process of this master thesis. Besides Marjolein, Merel and Sergey, I would also like to acknowledge a few academics whose research was of much help for the present thesis: these are Ramon Ferrer i Cancho, Barbara Köpke, Monika S. Schmid and Steven T. Piantadosi. And last but not least, I want to thank Gerrit Bloothooft for the time and interest to be the second reader of this thesis, and for being just as helpful in previous occasions during my master's program.

I would also like to thank the people who made my experience in the Utrecht University much more enjoyable. Georg for his infinite help and for the existentialist coffees. Timo and Lydia, for being there in the good times and the bad. Borja, Daniel, Estefanía, Gui, Mikey and Karim, for being like a second family here. And everyone I am not mentioning but know should be here. Without them Utrecht would not have been half as good.

Y cómo no, a papá y mamá, porque de no ser por vosotros y vuestro apoyo incesante desde mi mismo nacimiento, esto jamás hubiese sido posible.

References

Adamic, L. A., & Huberman, B. A. (2002). Zipf's law and the Internet. *Glottometrics*, 3(1), 143–150.

Almor, A., Kempler, D., MacDonald, M. C., Andersen, E. S., & Tyler, L. K. (1999). Why do Alzheimer patients have difficulty with pronouns? Working memory, semantics, and reference in comprehension and production in Alzheimer's disease. *Brain and Language*, 67(3), 202–227.

Ammerlaan, T. (1996). *You Get a Bit Wobbly...: Exploring Bilingual Lexical Retrieval Processes in the Context of First Language Attrition.* Unpublished doctoral thesis.

Andersen, R. W. (1982). Determining the linguistic attributes of language attrition. *The Loss of Language Skills*, 83, 118.

Aronoff, J. M., Gonnerman, L. M., Almor, A., Arunachalam, S., Kempler, D., & Andersen, E. S. (2006). Information content versus relational knowledge: semantic deficits in patients with Alzheimer's disease. *Neuropsychologia*, 44(1), 21–35.

Ashcraft, M. H. (1989). *Human memory and cognition.* Scott, Foresman & Co.

Avrutin, S. (2006). Weak syntax. *Broca's Region*, 49–62.

Avrutin, S., Haverkort, M., & Hout, A. van (2001). Language acquisition and language breakdown. *Brain and Language*, 77(3), 269–273.

Baayen, R. H. (2001). *Word frequency distributions* (Vol. 18). Springer Science & Business Media.

Bahrick, H. P. (1984). Fifty years of second language attrition: Implications for programmatic research. *The Modern Language Journal*, 68(2), 105–118.

Bailey, C.-J. N. (1973). The patterning of language variation. In R. W. Bailey & J. L. Robinson (Eds.), *Varieties of present day English*, 156 – 189. New York: MacMillian.

Baixeries, J., Elvevåag, B., & Ferrer-i-Cancho, R. (2013). The evolution of the exponent of Zipf's law in language ontogeny. *PloS One*, 8(3), e53227.

Baroni, M. (2009). Distributions in text. *Corpus Linguistics: An International Handbook*, 2, 803–821.

Barrera, A., McKenna, P. J., & Berrios, G. E. (2005). Formal thought disorder in schizophrenia: an executive or a semantic deficit?, *Psychological Medicine*, 35(01), 121–132.

Bastiaanse, R., & Bol, G. (2001). Verb inflection and verb diversity in three populations: Agrammatic speakers, normally developing children, and children with specific language impairment (SLI). *Brain and Language*, 77(3), 274–282.

Bot, K. de (2007). Dynamic systems theory, lifespan development and language attrition. *Language Attrition: Theoretical Perspectives*, 53–68.

Bot, K. de, & Clyne, M. (1989). Language reversion revisited. *Studies in Second Language Acquisition*, 11(02), 167–177.

Bot, K. de, & Clyne, M. (1994). A 16-year longitudinal study of language attrition in Dutch immigrants in Australia. *Journal of Multilingual & Multicultural Development*, 15(1), 17–28.

Bot, K. de, Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10(01), 7–21.

Bot, K. de, Martens, V., & Stoessel, S. (2004). Finding residual lexical knowledge: The "Savings" approach to testing vocabulary. *International Journal of Bilingualism*, 8(3), 373-382.

Burke, D. M. (2002). Language production and aging. In *Constraints on language: Aging, grammar, and memory*, 3–27. Springer.

Butterworth, B., Howard, D., & Mcloughlin, P. (1984). The semantic deficit in aphasia: The relationship between semantic errors in auditory comprehension and picture naming. *Neuropsychologia*, 22(4), 409–426.

Camicioli, R., Oken, B. S., Sexton, G., Kaye, J. A., & Nutt, J. G. (1998). Verbal fluency task affects gait in Parkinson's disease with motor freezing. *Journal of Geriatric Psychiatry and Neurology*, 11(4), 181–185.

Caramazza, A., & Berndt, R. S. (1978). Semantic and syntactic processes in aphasia: A review of the literature. *Psychological Bulletin*, 85(4), 898.

Carnero, C. & Lendínez, A. (1999). Utilidad del test de fluencia verbal semántica en el diagnóstico de la demencia. *Revista de Neurología*, 29, 709-714.

Carnero, C., Maestre, J., Marta, J., Mola, S., Olivares, J. & Sempere, A.P. (2000). Validación de un modelo de predicción de fluidez verbal semántica. *Revista de Neurología*, 30, 1012-1015.

Chater, N., & Brown, G. D. (1999). Scale-invariance as a unifying psychological principle. *Cognition*, 69(3), B17–B24.

Chen, Y.-S. (1991). Zipf's law in natural languages, programming languages, and command languages: the Simon-Yule approach. *International Journal of Systems Science*, 22(11), 2299–2312.

Cherciov, M. (2013). Investigating the impact of attitude on first language attrition and second language acquisition from a Dynamic Systems Theory perspective. *International Journal of Bilingualism*, 17(6), 716–733.

Chitashvili, R. J., & Baayen, R. H. (1993). Word frequency distributions. *Quantitative text analysis*, 54-135.

Christman, S. S. (2002). Dynamic systems theory: Application to language development and acquired aphasia. *Connectionist Approaches to Clinical Problems in Speech and Language: Therapeutic and Scientific Applications*, 111–146.

Conrad, B., & Mitzenmacher, M. (2004). Power laws for monkeys typing randomly: the case of unequal probabilities. *Information Theory*, IEEE Transactions on, 50(7), 1403–1414.

Cook, V. (2003). *Effects of the second language on the first* (Vol. 3). Multilingual matters.

Corominas-Murtra, B., Hanel, R., & Thurner, S. (2015). Understanding scaling through history-dependent processes with collapsing sample space. *Proceedings of the National Academy of Sciences*, 112(17), 5348–5353.

Corominas-Murtra, B., & Solé, R. V. (2010). Universality of Zipf's law. *Physical Review E*, 82(1), 011102.

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type-token ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2), 94–100.

Cuetos-Vega, F., Menéndez-González, M. & Calatayud-Noguera, T. (2007). Descripción de un nuevo test para la detección precoz de la enfermedad de Alzheimer. *Revista de Neurología*, 44(8), 469-474.

Dahlgren, D. J. (1998). Impact of knowledge and age on tip-of-the-tongue rates. *Experimental Aging Research*, 24(2), 139–153.

Dostert, S. (2009). *Multilingualism, L1 attrition and the concept of "native speaker."* Unpublished PhD Dissertation, Heinrich-Heine Universität, Düsseldorf.

Dressler, W. U. (1991). The sociolinguistic and patholinguistic attrition of Breton phonology, morphology, and morphonology. *First Language Attrition*, 99–112.

Du Bois, I. (2015). Language attrition and code-switching among US Americans in Germany. *Stellenbosch Papers in Linguistics Plus*, 39, 1–16.

Durán, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental Trends in Lexical Diversity. *Applied Linguistics*, 25(2), 220–242.

Ecke, P. (2004). Language attrition and theories of forgetting: A cross-disciplinary review. *International Journal of Bilingualism*, 8(3), 321–354.

Egmond, M. van (2011). *Word Finding Difficulties in Aphasia and their Effect on Zipf's Law*. Unpublished MA thesis. Universiteit Utrecht.

Egmond, M. van, Ewijk, L. van, & Avrutin, S. (2015). Zipf's Law in Non-Fluent Aphasia. *Journal of Quantitative Linguistics*, 22(3), 233–249.

Evrard, M. (2001). *Contribution à l'étude de l'accès lexical aux noms propres chez des adultes d'âges différents en situation de dénomination.* Atelier national de reproduction des thèses. Université Toulouse 2 Jean Jaurès.

Ewijk, L. van. (2013). *Word retrieval in acquired and developmental language disorders: A bit more on processing.* Netherlands Graduate School of Linguistics.

Farmer, J. D., & Geanakoplos, J. (2008). Power laws in economics and elsewhere. Santa Fe Institute. Yale University.

Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology*, 25(11), 1414–1430.

Ferrer i Cancho, R. (2005b). Decoding least effort and scaling in signal frequency distributions. *Physica A: Statistical Mechanics and Its Applications*, 345(1), 275–284.

Ferrer i Cancho, R. (2005a). The variation of Zipf's law in human language. *The European Physical Journal B-Condensed Matter and Complex Systems*, 44(2), 249–257.

Ferrer i Cancho, R. (2006). When language breaks into pieces. A conflict between communication through isolated signals and language. *Bio Systems*, 84(3), 242–253.

Ferrer i Cancho, R., Riordan, O., & Bollobás, B. (2005). The consequences of Zipf's law for syntax and symbolic reference. *Proceedings of the Royal Society of London B: Biological Sciences*, 272(1562), 561–565.

Ferrer i Cancho, R., & Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, 100(3), 788–791.

Fishman, J. A. (1991). *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages* (Vol. 76). Multilingual matters.

Font-Clos, F., Boleda, G., & Corral, Á. (2013). A scaling law beyond Zipf's law and its relation to Heaps' law. *New Journal of Physics*, 15(9), 093033.

Garcés, M., Santos, S., Pérez, C. & Pascual, L.F. (2004). Test del supermercado: datos normativos preliminares en nuestro medio. *Revista de Neurología*, 39, 415-418.

Garrard, P., Lambon Ralph, M. A., Patterson, K., Pratt, K. H., & Hodges, J. R. (2005). Semantic feature knowledge and picture naming in dementia of Alzheimer's type: A new approach. *Brain and Language*, 93(1), 79–94.

Gilvarry, C. M., Russell, A., Jones, P., Sham, P., Hemsley, D., & Murray, R. M. (2001). Verbal fluency in patients with schizophrenia and affective psychoses and their first-degree relatives. *Psychological Medicine*, 31(04), 695–704.

Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. Neuropsychology, 16(4), 562.

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *The assessment of aphasia and related disorders*. Lippincott Williams & Wilkins.

Goral, M. (2004). First-language decline in healthy aging: Implications for attrition in bilingualism. *Journal of Neurolinguistics*, 17(1), 31–52.

Grodzinsky, Y. (1990). *Theoretical perspectives on language deficits*. MIT press.

Grosjean, F. (1997). Processing mixed language: Issues, findings, and models. *Tutorials in Bilingualism: Psycholinguistic Perspectives*, 225–254.

Gross, S. (2004). A modest proposal: Explaining language attrition in the context. *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*, 28, 281.

Guiraud, P. (1971). The Semic Matrices of Meaning. *Essays in Semiotics*, Edited by J. Kristeva, J. Rey-Debove and DJ Umiker. The Hague and Paris: Mouton, 150–159.

Gürel, A. (2002). *Linguistic characteristics of second language acquisition and first language attrition: Turkish overt versus null pronouns*. McGill University, Montréal.

Gürel, A. (2004). Selectivity in L2-induced L1 attrition: a psycholinguistic account. *Journal of Neurolinguistics*, 17(1), 53–78.

Gürel, A. (2011). In search for a unified model of L2 acquisition and L1 attrition: A commentary for the Interface Hypothesis. *Linguistic Approaches to Bilingualism*, 1(1), 39–42.

Håkansson, G. (1995). Syntax and morphology in language attrition: A study of five bilingual expatriate Swedes. *International Journal of Applied Linguistics*, 5(2), 153–169.

Herdina, P., & Jessner, U. (2002). *A dynamic model of multilingualism: Perspectives of change in psycholinguistics* (Vol. 121). Multilingual Matters.

Hermans, S. P. F. (2012). Attrition and Language Reversion in Dutch-English Migrants in Australia: Working Memory Capacity and Inhibitory Functioning across Languages and Age Groups. Unpublished MA thesis. Utrecht University.

Hernández-Fernández, A., & Diéguez-Vide, F. (2013). Zipf's law and the detection of the verbal evolution in Alzheimer's disease. *Anuario de Psicología/The UB Journal of Psychology*, 43(1), 67–82.

Hier, D. B., Hagenlocker, K., & Shindler, A. G. (1985). Language disintegration in dementia: Effects of etiology and severity. *Brain and Language*, 25(1), 117–133.

Hinton, L., & Hale, K. (2001). *The green book of language revitalization in practice.* Brill.

Hulsen, M. E. H. (2000). *Language loss and language processing: Three generations of Dutch migrants in New Zealand.* [Sl: sn].

Hutz, M. (2004). Is there a natural process of decay? *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*, 28, 189.

Isurin, L. (2000). Deserted island or a child's first language forgetting. *Bilingualism: Language and Cognition*, 3(02), 151–166.

Izsák, J. (2006). Some practical aspects of fitting and testing the Zipf-Mandelbrot model: A short essay. *Scientometrics*, 67(1), 107-120.

Jakobson, R. (1941). *Kindersprache, Aphasie, und allgemeine Lautgesetze.* Uppsala, Almqvist & Wiksell. (Child Language, Aphasia and Phonological Universals) The Hague: Mouton.

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing*, 19(1), 57–84.

Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, 63(s1), 87–106.

Jessner, U. (2008). A DST model of multilingualism and the role of metalinguistic awareness. *The Modern Language Journal*, 92(2), 270–283.

Jia, G., & Aaronson, D. (1999). Age differences in second language acquisition: The dominant language switch and maintenance hypothesis. In *Proceedings of the 23rd annual Boston University conference on language development*, Vol. 23, 301–312.

Johnson, W. (1939). *Language and Speech Hygiene: An Application of General Semantics.* Outline of a Course. Institute of General Semantics.

Johnson, W. (1944). A program of research. *Psychological Monographs*, 56(2), 1.

Kapanga, A. (1998). Impact of language variation and accommodation theory on language maintenance: An analysis of Shaba Swahili. *Endangered Languages–current Issues and Future Prospects*, 261–288.

Kargar, H., & Rezai, M. J. (2014). The Contribution of Lexicon to the Language Attrition of Iranian Immigrants to Canada. *Procedia - Social and Behavioral Sciences*, 98, 830–838.

Kaufman, D. & Aronoff, M. (1991). Morphological disintegration and reconstruction in first language attrition. In H.W. Seliger & R. M. Vago (Eds.), *First Language Attrition*, 175-188. Cambridge: Cambridge University Press.

Kawamura, K., & Hatano, N. (2002). Universality of Zipf's law. *Journal of the Physical Society of Japan*, 71(5), 1211–1213.

Keijzer, M. (2004). First language attrition: A cross-linguistic investigation of Jakobson's regression hypothesis. *International Journal of Bilingualism*, 8(3), 389–393.

Keijzer, M. (2007). *Last in first out? An investigation of the regression hypothesis in Dutch emigrants in Anglophone Canada.* Netherlands Graduate School of Linguistics.

Keijzer, M. (2010). The regression hypothesis as a framework for first language attrition. *Bilingualism: Language and Cognition*, 13(01), 9–18.

Kello, C. T., Brown, G. D., Ferrer-i-Cancho, R., Holden, J. G., Linkenkaer-Hansen, K., Rhodes, T., & Orden, G. C. van (2010). Scaling laws in cognitive sciences. *Trends in Cognitive Sciences*, 14(5), 223–232.

Kempler, D., Curtiss, S., & Jackson, C. (1987). Syntactic preservation in Alzheimer's disease. *Journal of Speech, Language, and Hearing Research*, 30(3), 343–350.

Kohn, S. E., & Goodglass, H. (1985). Picture-naming in aphasia. *Brain and Language*, 24(2), 266–283.

Kolk, H. (2001). Does agrammatic speech constitute a regression to child language? A three-way comparison between agrammatic, child, and normal ellipsis. *Brain and Language*, 77(3), 340–350.

Köpke, B. (2002). Activation thresholds and non-pathological first language attrition. *Advances in the Neurolinguistics of Bilingualism*, 119–142.

Köpke, B. (2007). Language attrition at the crossroads of brain, mind, and society. *Language Attrition: Theoretical Perspectives*, 9–38.

Köpke, B., & Nespoulous, J. L. (2001). First language attrition in production skills and metalinguistic abilities in German-English and German-French bilinguals. *Sociolinguistic and Psycholinguistic Perspectives on Maintenance and Loss of Minority Languages*, 221–234.

Köpke, B., & Schmid, M. S. (2004). *Language attrition. First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*, 28(1).

Kroll, J. F. (1993). Accessing conceptual representations for words in a second language. *The Bilingual Lexicon*, 53481.

Kroll, J. F., & Tokowicz, N. (2001). The Development of Conceptual Representation for Words in a Second Language.

Kroll, J. F., Tokowicz, N., & others. (2005). Models of bilingual representation and processing. *Handbook of Bilingualism: Psycholinguistic Approaches*, 531–553.

Lahey, M., & Edwards, J. (1999). Naming errors of children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 42(1), 195–205.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, 18(2), 141–165.

Laufer, B., & Baladzhaeva, L. (2015). First language attrition without second language acquisition: An exploratory study. *ITL-International Journal of Applied Linguistics*, 166(2), 229–253.

Lee, J.-W., & Schallert, D. L. (1997). The Relative Contribution of L2 Language Proficiency and L1 Reading Ability to L2 Reading Performance: A Test of the Threshold Hypothesis in an EFL Context. *TESOL Quarterly*, 31(4), 713–739.

Levine, G. (1996). Elderly second-generation speakers of Yiddish: Toward a model of L1 loss, incomplete L1 acquisition, competence, and control. *Southwest Journal of Linguistics*, 15(1-2), 109–120.

Li, W. (1992). Random texts exhibit Zipf's-law-like word frequency distribution. *Information Theory*, IEEE Transactions on, 38(6), 1842–1845.

Li, W. (2002). Zipf's law everywhere. *Glottometrics*, 5, 14–21.

Louridas, P., Spinellis, D., & Vlachos, V. (2008). Power laws in software. *ACM Transactions on Software Engineering and Methodology* (TOSEM), 18(1), 2.

MacLeod, C. M. (1976). Bilingual episodic memory: Acquisition and forgetting. *Journal of Verbal Learning and Verbal Behavior*, 15(4), 347–364.

MacWhinney, B. (2000a). The CHILDES project: Tools for analyzing talk: Volume I: Transcription format and programs, volume II: The database. *Computational Linguistics*, 26(4), 657–657.

MacWhinney, B. (2000b). The CHILDES project: The database (Vol. 2). *Psychology Press.*

MacWhinney, B. (1998). Models of the emergence of language. *Annual Review of Psychology*, 49(1), 199–227.

Malvern, D. D., Richards, B. J., Chipere, N., & Duran, P. (2004). *Lexical diversity and language development.* Houndmills, Hampshire, UK: Palgrave Macmillan.

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 19(1), 85–104.

Malvern, D., & Richards, B. (2012). Measures of Lexical Richness. In *The Encyclopedia of Applied Linguistics.* Blackwell Publishing Ltd..

Manaris, B., Romero, J., Machado, P., Krehbiel, D., Hirzel, T., Pharr, W., & Davis, R. B. (2005). Zipf's law, music classification, and aesthetics. *Computer Music Journal*, 29(1), 55–69.

Mandelbrot, B. (1953). An informational theory of the statistical structure of language. *Communication Theory*, 84, 486–502.

Mandelbrot, B. (1961). On the theory of word frequencies and on related Markovian models of discourse. *Structure of Language and Its Mathematical Aspects*, 12, 190–219.

Manin, D. Y. (2008). Zipf's law and avoidance of excessive synonymy. *Cognitive Science*, 32(7), 1075–1098.

McCarthy, P. M., & Jarvis, S. (2007). VOCD: A theoretical and empirical evaluation. *Language Testing*, 24(4), 459–488.

McElree, B., Jia, G., & Litvak, A. (2000). The time course of conceptual processing in three bilingual populations. *Journal of Memory and Language*, 42(2), 229–254.

Meara, P. (2004). Modelling vocabulary loss. *Applied Linguistics*, 25(2), 137–155.

Michael, E. B., & Gollan, T. (2003). Being and becoming bilingual: Individual differences and consequences for language production. *Handbook of bilingualism: Psycholinguistic approaches.*

Miller, G. A. (1957). Some effects of intermittent silence. *The American Journal of Psychology*, 70(2), 311–314.

Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2), 226–251.

Montrul, S. (2002). Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition*, 5(01), 39–68.

Montrul, S. A. (2008). *Incomplete acquisition in bilingualism: Re-examining the age factor* (Vol. 39). John Benjamins Publishing.

Murphy, L. (2015). likelihood: Tools for Maximum Likelihood Estimation. R package version 1.7.

Naeser, M. A., Martin, P. I., Nicholas, M., Baker, E. H., Seekins, H., Kobayashi, M., … Pascual-Leone, A. (2005). Improved picture naming in chronic aphasia after TMS to part of right Broca's area: An open-protocol study. *Brain and Language*, 93(1), 95–105.

Neisser, U. (1984). Interpreting Harry Bahrick's discovery: What confers immunity against forgetting? *Journal of Experimental Psychology: General*, Vol 113(1), 32-35.

Nicoladis, E., & Grabois, H. (2002). Learning English and losing Chinese: A case study of a child adopted from China. *International Journal of Bilingualism*, 6(4), 441–454.

Nishimura, Y., Takahashi, K., Ohtani, T., Ikeda-Sugita, R., Kasai, K., & Okazaki, Y. (2015). Dorsolateral prefrontal hemodynamic responses during a verbal fluency task in hypomanic bipolar disorder. *Bipolar Disorders*, 17(2), 172–183.

Okada, G., Okamoto, Y., Morinobu, S., Yamawaki, S., & Yokota, N. (2003). Attenuated left prefrontal activation during a verbal fluency task in patients with depression. *Neuropsychobiology*, 47(1), 21–26.

Olshtain, E. (1989). Is second language attrition the reversal of second language acquisition? *Studies in Second Language Acquisition*, 11(02), 151–165.

Opitz, C. (2011). *First language attrition and second language acquisition in a second language environment*. Ph.D. dissertation. Trinity College Dublin.

Pan, B. A., & Gleason, J. B. (1986). The study of language loss: Models and hypotheses for an emerging discipline. *Applied Psycholinguistics*, 7(03), 193–206.

Paradis, M. (1985). On the representation of two languages in one brain. *Language Sciences*, 7(1), 1–39.

Paradis, M. (1993). Linguistic, psycholinguistic, and neurolinguistic aspects of "interference" in bilingual speakers: The activation threshold hypothesis. *International Journal of Psycholinguistics*, Vol 9(2), 133-145.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism* (Vol. 18). John Benjamins Publishing.

Paradis, M. (2007). L1 attrition features predicted by a neurolinguistic theory of bilingualism. *Language Attrition: Theoretical Perspectives*, 121–133.

Parker-Rhodes, A. F., & Joyce, T. (1956). A theory of word-frequency distribution. *Nature*, 178, 1308.

Patterson, K. E., Graham, N., & Hodges, J. R. (1994). Reading in dementia of the Alzheimer type: A preserved ability? *Neuropsychology*, 8(3), 395–407.

Pavlenko, A. (2004). L2 influence and L1 attrition in adult bilingualism. *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*, 47–59.

Pavlenko, A. (2010). Verbs of motion in L1 Russian of Russian–English bilinguals. *Bilingualism: Language and Cognition*, 13(01), 49–62.

Pavlenko, A., & Jarvis, S. (2002). Bidirectional transfer. *Applied Linguistics*, 23(2), 190–214.

Pease, D. M., & Goodglass, H. (1978). The Effects of Cuing on Picture Naming in Aphasia. *Cortex*, 14(2), 178–189.

Pekkala, S., Wiener, D., Himali, J. J., Beiser, A. S., Obler, L. K., Liu, Y., ... others. (2013). Lexical retrieval in discourse: An early indicator of Alzheimer's dementia. *Clinical Linguistics & Phonetics*, 27(12), 905–921.

Pfaff, C. W. (1981). Sociolinguistic problems of immigrants: foreign workers and their children in Germany (a review article). *Language in Society*, 10(02), 155–188.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, 21(5), 1112–1130.

Piotrowska, W., & Piotrowska, X. (2004). Pathological text and its statistical parameters. *Journal of Quantitative Linguistics*, 11(2), 133–140.

Piotrowski, R. G., & Spivak, D. L. (2007). Linguistic disorders and pathologies: synergetic aspects. *Exact Methods in the Study of Language and Text. To Honor Gabriel Altmann*, Berlin: Gruyter, 545–554.

Polinsky, M. (2006). Incomplete acquisition: American russian. *Journal of Slavic Linguistics*, 14(2), 191–262.

Py, B. (1986). Native language attrition amongst migrant workers: Towards an extension of the concept of interlanguage. *Crosslinguistic Influence in Second Language Acquisition*, 163–172.

Quan, W., Wu, T., Li, Z., Wang, Y., Dong, W., & Lv, B. (2015). Reduced prefrontal activation during a verbal fluency task in Chinese-speaking patients with schizophrenia as measured by near-infrared spectroscopy. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 58, 51–58.

Ribot, T. A. (1882). *Diseases of memory, an essay in the positive psychology* (Vol. 43). Kegan Paul, Trench.

Roberts, P. M., & Le Dorze, G. (1998). Bilingual aphasia: Semantic organization, strategy use, and productivity in semantic verbal fluency. *Brain and Language*, 65(2), 287–312.

Sasse, H.-J. (1992). Language decay and contact-induced change: Similarities and differences. *Contributions to the Sociology of Language*, 64, 59–59.

Saville-Troike, M., Pan, J., & Dutkova, L. (1995). Differential Effects of L2 on Children's L1 Development/Attrition. *Southwest Journal of Linguistics*, 14, 125–49.

Schmid, M. S. (2002). *First language attrition, use and maintenance: The case of German Jews in Anglophone countries* (Vol. 24). John Benjamins Publishing.

Schmid, M. S. (2006). Problems of measurement in language attrition research. Paper presented at the workshop "Measurement of bilingual proficiency" at the 16th Sociolinguistics Symposium, Limerick, July 2006.

Schmid, M. S. (2011). *Language attrition*. Cambridge University Press.

Schmid, M. S. (2012). The impact of age and exposure on bilingual development in international adoptees and family migrants: A perspective from Holocaust survivors. *Linguistic Approaches to Bilingualism*, 2(2), 177–208.

Schmid, M. S. (2013). First language attrition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(2), 117–123.

Schmid, M. S., & Jarvis, S. (2014). Lexical access and lexical diversity in first language attrition. *Bilingualism*, 17(4), 729.

Schmid, M. S., & Keijzer, M. (2009). First language attrition and reversion among older migrants. *International Journal of the Sociology of Language*, 2009(200), 83–101.

Schmid, M. S., & Köpke, B. (2007). Bilingualism and attrition. Language Attrition: Theoretical Perspectives. In: *Language Attrition: Theoretical Perspectives* (Vol. 33). John Benjamins Publishing.

Schmid, M. S., & Köpke, B. (2008). L1 attrition and the mental lexicon. In: Pavlenko, A. (ed.) *The Bilingual Mental Lexicon: interdisciplinary approaches*, 209–238. Clevedon: Multilingual Matters.

Schmid, M. S., & Köpke, B. (2009). L1 attrition and the mental lexicon. *The Bilingual Mental Lexicon: Interdisciplinary Approaches*, 209–238.

Schoenmakers-Klein Gunnewiek, M. (1998). *Taalverlies door taalcontact? Een onderzoek bij Portugese migranten.* Doctoral dissertation, Katholieke Universiteit Brabant.

Segalowitz, N. (1991). Does advanced skill in a second language reduce automaticity in the first language? *Language Learning*, 41(1), 59–83.

Seliger, H. W., & Vago, R. M. (1991). *First language attrition.* Cambridge University Press.

Shuttleworth, E. C., & Huber, S. J. (1988). The naming disorder of dementia of Alzheimer type. *Brain and Language*, 34(2), 222–234.

Slobin, D. I. (1975). *Language change in childhood and in history.* Language Behavior Research Laboratory, University of California.

Smith, R. D. (2008). Investigation of the Zipf-plot of the extinct Meroitic language. *Glottometrics* 15, 2007, 53-61.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2), 174.

Soesman, A. (1997). An experimental study on native language attrition in Dutch adult immigrants in Israel. *Dutch Overseas: Studies in Maintenance and Loss of Dutch as an Immigrant Language*, 181–194.

Sorace, A. (2005). Selective optionality in language development. *Syntax and Variation: Reconciling the Biological and the Social*, 55–80.

Spear, N. E., & Riccio, D. C. (1994). *Memory: Phenomena and principles.* Allyn & Bacon.

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

Swan, D., & Goswami, U. (1997). Picture naming deficits in developmental dyslexia: The phonological representations hypothesis. *Brain and Language*, 56(3), 334–353.

Swinney, D., Prather, P., & Love, T. (2000). The time-course of lexical access and the role of context: Converging evidence from normal and aphasic processing. *Language and the Brain: Representation and Processing*, 273–292.

Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*. MIT press.

Thorndike, E. L. (1913). *The psychology of learning* (Vol. 2). Teachers College, Columbia University.

Thurner, S., Hanel, R., Liu, B., & Corominas-Murtra, B. (2015). Understanding Zipf's law of word frequencies through sample-space collapse in sentence formation. *Journal of The Royal Society Interface*, 12(108), 20150330.

Tomiyama, M. (2000). Child second language attrition: A longitudinal case study. *Applied Linguistics*, 21(3), 304–332.

Tsimpli, I. (2007). First language attrition from a minimalist perspective. *Language Attrition: Theoretical Perspectives*, 33, 83.

Tsimpli, I., Sorace, A., Heycock, C., & Filiaci, F. (2004). First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism*, 8(3), 257–277.

Utman, J. A., Blumstein, S. E., & Sullivan, K. (2001). Mapping from sound to meaning: Reduced lexical activation in Broca's aphasics. *Brain and Language*, 79(3), 444–472.

Vakhtin, N. (1998). Copper Island Aleut: A case of language 'resurrection.' In L. A. Grenoble & L. J. Whaley (Eds.), *Endangered languages: Language loss and community response*, 317–327. Cambridge, U.K.: Cambridge University Press.

Verhoeven, L., & Balkom, H. van (2003). *Classification of developmental language disorders: Theoretical issues and clinical implications*. Psychology Press.

Waas, M. (1996). *Language attrition downunder*. Frankfurt: Peter Lang.

Wachal, R. S., & Spreen, O. (1973). Some Measures of Lexical Diversity in Aphasic and Normal Language Performance. *Language and Speech*, 16(2), 169–181.

Weinreich, U. (1979). *Languages in contact: Findings and problems*. Walter de Gruyter.

Weltens, B., Grendel, M., Schreuder, R., & Weltens, B. (1993). Attrition of vocabulary knowledge. *The Bilingual Lexicon*, 135–156.

Weltens, B., Els, T. J. van, & Schils, E. (1989). The long-term retention of French by Dutch students. *Studies in Second Language Acquisition*, 11(2), 205–216.

Wickelgren, W. A. (1977). *Learning and memory.* Prentice-Hall.

Wilde, O. (1994). *The Picture of Dorian Gray.* Retrieved from the Gutenberg project (www.gutenberg.org). E-book number 174.

Wixted, J. T. (2004b). On Common Ground: Jost's (1897) law of forgetting and Ribot's (1881) law of retrograde amnesia. *Psychological Review*, 111(4), 864.

Wixted, J. T. (2004a). The psychology and neuroscience of forgetting. *Annu. Rev. Psychol.*, 55, 235–269.

Wixted, J. T., & Ebbesen, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & Cognition*, 25(5), 731–739.

Wright, H. H., Silverman, S., & Newhoff, M. (2003). Measures of lexical diversity in aphasia. *Aphasiology*, 17(5), 443–452.

Yağmur, K. (1997). *First language attrition among Turkish speakers in Sydney.* PhD Dissertation. Tilburg University Press.

Yee, E., Blumstein, S. E., & Sedivy, J. C. (2008). Lexical-semantic activation in Broca's and Wernicke's aphasia: Evidence from eye movements. *Journal of Cognitive Neuroscience*, 20(4), 592–612.

Yeni-Komshian, G. H., Flege, J. E., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean–English bilinguals. *Bilingualism: Language and Cognition*, 3(02), 131–149.

Yilmaz, G., & Schmid, M. S. (2012). L1 accessibility among Turkish-Dutch bilinguals. *The Mental Lexicon*, 7(3), 249–274.

Yukawa, E. (1997). *L1 Japanese attrition and regaining: Three case studies of two early bilingual children.* PhD Dissertation. Centre for Research on Bilingualism, Stockholm University.

Zelealem, L. (2000). *The Kemantney language: A sociolinguistic and grammatical study of language replacement.* PhD Dissertation. Addis Adaba University.

Zipf, G. K. (1935). *The psycho-biology of language.* Oxford, England: Houghton, Mifflin.

Zipf, G. K. (1949). Human behavior and the principle of least effort. *The Southwestern Social Science Quarterly* Vol. 30, No. 2, 147-149

List of Figures

List of Tables