

Grounded Theory Evaluations of Second Language Vocabulary Acquisition

Grandon Goertz¹, Terese Anderson²

¹University of New Mexico – sordfish@unm.edu

²University of Chicago – B13@uchicago.edu

Abstract

There are numerous theories of second language vocabulary acquisition and teaching. This research employs a mixed methods with a grounded approach to determine how second language students learn new vocabulary and to propose a theory of vocabulary learning. Students attending the total immersion German Deutsche Summerschule (June 2019) logged 1,514 new vocabulary words. Word frequencies were determined by comparisons to available textbooks and to five German linguistic corpora. After plotting vocabulary frequencies and comparing the results to a variety of mathematical explanations, it was found that learning could be explained by specific mathematical equations, showing that vocabulary memory was constructed in the form of an expanding neural network, with commonly used words representing nodes being connected to words and collocations.

1 Introduction

Language immersion programs are often cited as examples of efficient second language learning. Immersion programs have been accepted as one of the most efficient methods of teaching a foreign language, and to that end some university foreign language departments sponsor study-abroad programs in host countries. University-level total immersion programs typically place second language students a country where the study language is natively spoken. Students study

abroad for a specific period of time, such as summer months, or a semester. Students generally do improve their second language (L2) vocabulary, but more, they improve their confidence in speaking and grammar use through continual, L2 exposure and language practice.

This mixed methods research study is designed to investigate the specific method and mechanism of L2 vocabulary acquisition in a total-immersion German language summer program.

1.1 Total immersion vocabulary acquisition

Students generally show language improvement after attending a study abroad or immersion program. DeKeyser (2007) reports that large fluency gains in some study abroad programs have happened. Incidental and repeated vocabulary learning is important to L2 vocabulary learning (Hunt & Beglar, 2005; Laufer, 2001; Nation, 2001, 2008; Nation & Webb, 2011; Schmitt, 2000, 2008; Webb & Chang, 2012). Immersion programs generally provide a level of learning that is more comprehensive than classroom language learning, partly due to the continuity of the program and the greatly increased use of the L2.

Student's writing exercises in immersion programs have been studied for the changes in performance. Writing is the easiest data to quantify because the work product can be collected and analyzed. Writing evaluations can be conducted from the perspectives of vocabulary use, grammar, accuracy, and vocabulary development. However, research results show variations in results. Freed, So, and Lazar (2003) found no significant progress in written fluency, and Pérez-Vidal and Juan-Garau (2009) found that for advanced instructed learners in a study abroad program slight improvements could be seen the lexical complexity of writing. In contrast, Meara (1995) found that vocabulary size increases about five times when a student studies in a foreign country. Laufer (2003) found that

tasks such as writing and composition exercises led to more word acquisition than reading tasks alone.

2 Literature review

2.1 Vocabulary learning and textbook dependence

Vocabulary is often taught in the L2 classroom by having students learn lists of words and then later, the students are tested in various ways if they can accurately recall the translated meaning of those words. The testing practice fits the textbook paradigms of learning and instruction. This approach has the advantage of measurability in that testing will produce numbers that can be statistically evaluated.

Ellis and Wulff (2015) observe that most L2 curriculum and instruction is driven by L2 textbooks which present the target language in terms of rules and vocabulary items. This observation agrees with those of Pellicer-Sánchez & Boers, 'The vast majority of studies empirically testing the efficiency of particular activities for vocabulary learning have focused on the acquisition of single-word items' (2018, p.154). A meta-analysis of 24 primary studies on L2 vocabulary learning showed that most research studies used single vocabulary words to test learning accomplishment (Webb, Uchihara & Yanagisawa, 2023). Their research review found that the most common way to test a student's L2 vocabulary acquisition has been through various testing methods such as, matching exercises; fill in the blanks exercises, completion drills, and vocabulary word definition drills. Archard (2018) and Lightbown (2008) indicate linguistic forms learned in isolation, such as vocabulary lists, are generally not available for use in communicative interaction.

2.2 Vocabulary acquisition according to Usage-based Theory

Language is described as a dynamic system (Larsen-Freeman, 1997; MacWhinney, 1999; Bybee & Hopper 2001; de Bot, Lowie, & Verspoor, 2007; Ellis, 2007a, 2008; Ellis & Larsen Freeman, 2009) which is subject to frequency effects of repeated words and collocations. Users have an extensive implicit knowledge of language sequences (Ellis, 2002) that include understandable collocations, idioms, and words commonly used together (Seidenberg & MacDonald, 1999; Christiansen & Chater, 2001; Ellis, 2008).

Functional approaches stress the role of context in vocabulary acquisition. Learners acquire vocabulary and the associated meanings through repeated exposures in various contexts. Vocabulary is seen as understanding the word meaning and how words combine with other words to form grammatical units.

Functional linguists advocate that vocabulary development is linked to meaning and word functioning and that vocabulary development is based on how words are used to convey ideas, and not by memorizing vocabulary lists. Vocabulary development is a dynamic process, driven by meaning and the socio-cultural environment.

An essential concept for Usage-based approaches is that language knowledge can be described as a system of prototypes, which have central prototype members and other related members that are part of the category, but less typical. L2 learners categorize new words and make connections between them (Boers, 2013) adding the new words to the appropriate prototypes.

2.3 Vocabulary acquisition according to Cognitive theory

Cognitive linguists state that the mental lexicon is not a long list of words, but words are organized into categories based on similarities and relationships. The lexicon is not a list of definitions, but is a dynamic cognitive network where words

are interconnected through various semantic, phonological, meanings, forms, and syntactic relationships (Li & Wang, 2024).

Cognitive linguistics claims that language acquisition is more than memorizing isolated words but also is about understanding how language reflects thought and how concepts are structured in the mind (Steen, 2018; Kohl, 2007). Cognitive linguistics considers that vocabulary learning develops through the process of language experiences and language use in context. Context plays a vital role in helping learners grasp the meaning.

Schema Theory explains that individuals use mental frameworks or schema to organize and interpret new information, such as new vocabulary. Mental frameworks help L2 learners make sense of new vocabulary by relating it to existing knowledge. In cases of contextualized learning, new words are not introduced in isolation but within relevant scenarios, and learners build stronger mental associations between words and their practical applications learning (Chamot, 2005; Wenden, 1998; Haukås, Bjørke, & Dypedahl, 2018; Hashemian & Nehzad, 2007).

2.4 Grounded approach: Method and purpose

Grounded methodology can offer new perspectives to old problems that remain unresolved after quantitative testing has been completed. Grounded theory is an approach that transcends paradigms and assumptions and attempts to develop new understandings of social processes (Conrad, 1982, p. 240; Maxwell, 2013; Miles, Huberman, & Saldaña, 2014, Friedman, 2023).

Grounded theory can produce theories that more closely explain the social reality of the classroom (Richter, 1975). The object of theory generation is to offer a new perspective on a situation and shared experiences that can be tested by other research methods. 'Thus, qualitative research should not be viewed as antagonistic toward or incompatible with quantitative methodologies. Qualitative

inquiry is a necessary and useful precursor to quantitative work' (Hutchinson, 2005, p. 132).

Grounded theory research is used to develop new understandings of social processes and to develop new theories. 'Grounded theory offers a perspective that reality is socially and symbolically constructed, always emerging and relative to the other facts of social life' (Hutchinson, 2005). Grounded theory is designed to discover or construct theory from data that is systematically obtained and analyzed using comparative analysis techniques (Glaser and Strauss, 1967; Richter, 1975; Conrad, 1982; Tie, Birks & Francis, 2019). Grounded theory research should explain a process relating to a particular phenomenon and the results should be communicated as a set of concepts that are related to each other.

Strength of grounded research comes from the fact that it employs a self-correction process. 'If a textualist shows their work and follows the correct order of operations, they are forced to grapple with counterarguments and they reduce the chance that they arrive at their conclusion because of priming and implicit bias. They are also less likely to engage in post-hoc rationalizations or backwards reasoning (Lee, 2021).' Data outliers do not need to be explained away or be removed by showing that they are on the ends of a normal distribution curve. The grounded approach allows the researcher to consider that one or more related phenomena can explain all the results (Loewen, Tuzcu, & Philp, 2023). Letting the data speak for itself helps avoid research bias in which the data is modified to fit a model, or in which the research answers are shaped by the research approach.

Grounded research employs data coding which is a cyclical process that is designed to show patterns of meaning through data relationships (Conrad, 1982). Coding identifies underlying patterns in the data and helps form ideas about relationships and properties of the data (Glaser & Strauss, 1967; Conrad, 1982; Miles, Huberman, & Saldaña, 2014).

3 Methods

This research approach does not aim to validate any particular theory but instead is designed to produce explanations and theories of a naturally occurring phenomenon.

3.1 Study population and environment

German language students attend the UNM *Deutsche Sommerschule* (German Summer School) with the goals of improving their German language proficiency, vocabulary development, and studying German culture. This language program is an annual event, and this study was conducted in June 2019. At this summer session, 23 students from various universities and 5 non-traditional students attended. Seventeen students participated in this research survey.

This immersion L2 environment is different from the classroom L2. The learning time is expanded from three hours a week in a traditional university class to 84 hours of uninterrupted hours per week in a variety of learning events.

The heavy emphasis on instruction and non-classroom activities in German provides an ideal environment to hear and use new vocabulary. Students complete the Summerschule with an enhanced vocabulary ability that was driven by formal study and numerous informal interactions that promote language learning.

3.2 Research questions

This study uses grounded theory to create a theory of vocabulary and language learning that is supported by the data. Research questions guide the formation of the theory, the objective of the research effort. Therefore, for this grounded approach, several research questions were developed:

1. What grammatical properties or patterns account for vocabulary learning?

2. What is a major process of vocabulary acquisition?

3.3 Data collection

At the beginning of the month-long language program, a brief presentation was made to the combined student assembly about the research project. Students were told that participation was entirely voluntary and confidential. Students were given small notebooks and asked to write, if they wanted to, new vocabulary words that they hear and learn.

The instructions included a comment that there were no right or wrong answers, and that words should be new and salient, including words that they find memorable, words that are used in ways that they have not heard before, and words that they would like to remember. Therefore, each student produced a unique list of vocabulary words.

Other data collection methods included researcher-participant observations, note taking, concept mapping, discussions with professors, and collecting the lists of new words the students provided. Data also included collected artifacts such as pertinent library materials, speeches, letters, and other unobtrusive sources of data. The teaching professors and staff, provided class handouts, files, and papers. Student notebooks were collected on the last day of the summer program.

3.4 Overview of the raw vocabulary data that was collected

At the end of the month-long session, the participating students provided 1,514 vocabulary words. A list of each student's vocabulary words was compiled, and the individual vocabulary lists were then compiled into one long list.

3.5 Data coding

Theory development began at the time of the completion of the program by making comparisons to the daily observations and coding sheets to the vocabulary lists. Relationships between concepts were drawn on coding maps to show the logical relationships between linguistic categories and the data. The process involved a succession of evaluations and tentative theories that were modified by evaluation methods. The vocabulary words were examined for patterns in the grammar type, part of speech, affixation, compound structure, semantic similarity, and other linguistic properties.

4 Discussion

Results are presented which respond to the research questions.

4.1 Research question number 1

The first research question asks, what grammatical properties or patterns would account for vocabulary learning? This question was aimed at determining if grammatical categories had an influence in vocabulary learning.

The 1,514 words were examined for similarities and qualities that would indicate that a particular method or common technique was being used by students to learn vocabulary. The distribution of prefixes (519 instances), compound words (186 instances), nouns, adjectives, and other grammatical classes and semantic values did not indicate that vocabulary listing was driven by grammatical type. The grammatical analysis does not offer an answer as to why students are identifying vocabulary words as salient, and the research did not find any connection between grammar and vocabulary saliency.

4.2 Research question number 2

The second research question asks: What is a major process of vocabulary acquisition? The solution to research question 2 requires inductive reasoning and outside research, which is encouraged by the Grounded theory research method. Inductive reasoning involves repetitive coding and an examination of the data.

4.2.1 Data property: Collocations

Students recorded vocabulary words, but also 152 collocations. Collocation identification is an indication of how intensely the students were listening to the content of conversations and lectures. Students expect to hear word combinations that are familiar and can be assimilated into the vocabulary. Collocation recognition indicates a general knowledge of word use combinations and a separation from word-by-word translations in phrases. Also, students were hearing collocations as word units that should be learned, such as idioms.

Collocations are salient because they have been said and remembered in the class environment, because they are new, and because students are developing an idea of the importance of collocations in expressing ideas efficiently. These results show that students are aware of the advantage of being able to use pre-made verbal chunks in discourse as a way of avoiding having to assemble sentences word by word.

Students were recording vocabulary in a functional-language environment taking noting individual words, collocations, and compounds. The fact that students were noting collocations in their word lists is also consistent with Ellis (2002, p.168), 'Constructions are independently represented units in a speaker's mind.' The fact that students intuitively identified the functional characteristics of the language environment is consistent with Weinert 1995; Nattinger & de Carrico 1992; Moon 1998, Howarth 1998; Wray 2000; Wray 2002; and Taylor & Littlemore 2014.

4.2.2 What explains the listing of high frequency words?

Some very high frequency words were listed by the L2 students. For this study, 107 student-provided vocabulary words were defined as being highly frequent, based on their membership in the German language Sketch Engine 13 corpus (de TenTen13, 16,526,335,416 words).

The important observation is that the students were hearing these words and writing them down. However, some of the more advanced students also recorded some of these frequent words, and the conclusion that can be drawn is that these frequent words were part of a new construction or phrase that had a different meaning from the definitions taught in lower class levels.

4.2.3 Middle frequent words

The majority of the words that were listed by the students would be considered moderately frequent, as defined on the plot of word frequencies with values set at a range of 2 words per million to 75 words per million based on the Sketch Engine 13 corpus.

4.2.4 Low frequency words

Students recorded 597 words that have a very low frequency which is defined as less than 100 instances per million words as identified in the Sketch Engine corpus.

In context, the conversations or class lectures were interesting and meaningful, and the students selected from the utterances the salient words that they did not know. A string of unintelligible words would not be meaningful to a student, and so no new vocabulary words would have been listed. The key factor here is that the students are picking words from a familiar word environment. Students did not listen simply for words they did not know, but rather they listened for words that they could discern from their use in context. A sentence

that was not understood would not transfer any meaning, and the student would not have written any vocabulary word down.

4.3 Mathematical implications

The vocabulary data has been identified as occurring in high, medium, and low frequency distributions, which indicates a mathematical explanation would be suited to explain vocabulary acquisition. This type of explanation would be consistent with Cognitive linguistic explanations of the structure of schemas and frameworks.

The L2 language learners in this study are engaged in a process of building a dynamic system of vocabulary and grammar through repetition and identification of words, collocations, and compounds as separate lexical units.

The key to answering this question lies in the ability to convert vocabulary words into frequency values for comparison to the total text vocabulary.

4.3.1 Textbook sources for vocabulary

It could be argued that students are simply extracting new vocabulary words from the textbooks and class materials they are using. However, of the 1,514 words that students identified as salient, 597 were not found in any of the course textbooks.

Four textbooks and four answer booklets, listed in the references, were used in the *Summerschule* for the formal class instruction. An inventory of all the words used in each textbook and each answer booklet was created, making eight individual word lists. The eight text lists were compiled into a Text Matrix with eight columns for comparison to the word list of the students' vocabulary words.

Other text information was gathered including the student newspaper, information postings, information sheets, PowerPoint presentations, and handouts from the lectures. These text sources were compiled into a single word list and

added to the Text Matrix, producing another larger matrix, the Combined Word Matrix. The Combined Word Matrix provides vocabulary that that students could be exposed to by reading, class activities, doing homework and social activities.

4.3.2 Coding the text words

A coding sheet was created to analyze the word distributions found in the Text Matrix. The word count of each word token was recorded, rank-ordered, and a histogram was plotted indicating a power-law mathematical function describes the word frequency distribution. Using Matlab's Data Analysis Toolbox, it was determined the distribution had a specific equation and a R^2 of 0.9687 goodness of fit.

Next, the eight Text Matrix frequency values were rank ordered, and the histograms were plotted, producing a characteristic power-law distribution functions with strong goodness of fit correlations.

The student vocabulary lists were found to collectively and individually be power-law distributed, reflecting the word distributions in the study text materials.

Table 1. The R^2 values for the power-law curves that describe the word frequency data for each textbook or booklet.

Source	R^2 score	Source	R^2 score
B1 Textbook	0.9739	B1 Answer book	0.9565
B2 Textbook	0.9584	B2 Answer book	0.9692
C1 Textbook	0.9782	C1 Answer book	0.9705
C2 Textbook	0.9585	C2 Answer book	0.9410

4.3.3 The logic of mathematical data evaluations: Power-law functions

A first step in evaluating data is to determine if the data can be represented by a function. By converting the individual vocabulary words to their frequency values, it is possible to plot their occurrences using a histogram.

The addition of vocabulary by the German language students was a continuous growth process, and this insight calls for the evaluation of the data to determine if the data could be represented by linear functions, exponential functions, hyperbolic functions (Eigen & Winkler, 1981), log-normal functions, or power-law functions (Clauset, Shalizi, & Newman, 2009). The data is most accurately represented as a power-law function of change.

Power-law distributions show up in a wide variety of natural and man-made phenomena, earth sciences, biology, ecology, paleontology, written and spoken texts, and social sciences (Clauset, Shalizi, & Newman, 2009). Power-law functions are common in many aspects of nature, biological growth, metabolic processes, and learning behavior.

The general formula for a power-law distribution is $y = ax^{-k}$, where a and $-k$ are constants, and y are the function values. The variable k usually approximates a simple multiple of $\frac{1}{4}$ (West, Brown, & Enquist, 2000), a property which was identified in this study data.

Power-law knowledge is a powerful tool to analyze data and to describe the characteristics of data, in this case, word frequencies and learning. Power functions grow by preferential attachment which is a process by which some quantity is added to a base according to how much is already there (Capocci, et al., 2006). For linguistic purposes, it means that vocabulary words tend to be added to a person's memory if there are related words that occupy the semantic or phonetic neighborhood.

4.3.4 Coding for word frequency

The next step in evaluating the vocabulary data is to compare the data to open-source and purchased German language corpora as a tool in the frequency coding process. Corpora are used to determine the frequency of the student vocabulary words. Each word in a corpus is given a numerical value of its frequency, which is the normalized frequency of occurrence. Plotting the corpus frequency of individual student words is a way of examining the data and determining if there is a pattern in the vocabulary acquisition.

A corpora analysis is used to determine the nature of the total language data and to determine if the plot of acquired vocabulary follows the mathematical shape of the total language data, or if other factors might be influencing vocabulary acquisition. Corpora comparisons provide the exact frequency distributions of the word.

Four corpora were chosen to evaluate the student vocabulary words: the Sketch Engine Corpus, the Falko Corpus, the *Datenband für Gesprochenes Deutsch* corpus, and the Leipzig Corpora. Specific corpora are used for specific research intents and these three showed the most likelihood of containing natural speech and text that would resemble the text material and the total educational environment. Corpora have the advantage of providing the scaled frequencies for words as they are used, and these corpora function as a weighted-use lexicon.

For example, Figure 1, below, shows the relative distribution of the 2000 most commonly used words in the total DGD corpus collection. Notice that the rank-order plot is almost a perfect power law curve, which is what would be expected. The power-law equation is shown and the R^2 value of 0.9921 indicates the data fit. This type of distribution shows that the use of the most common vocabulary words is according to the power-law principles of frequency. Relative frequency is shown on the y-axis, and the numerical count of the word is shown on the x-axis.

Figure 1. Plot of word frequencies of the 2,000 most frequent words from the DGD Corpus. Some of the vocabulary words are listed on the y-axis, but there is not sufficient space to display all the words. All corpora show similar results.

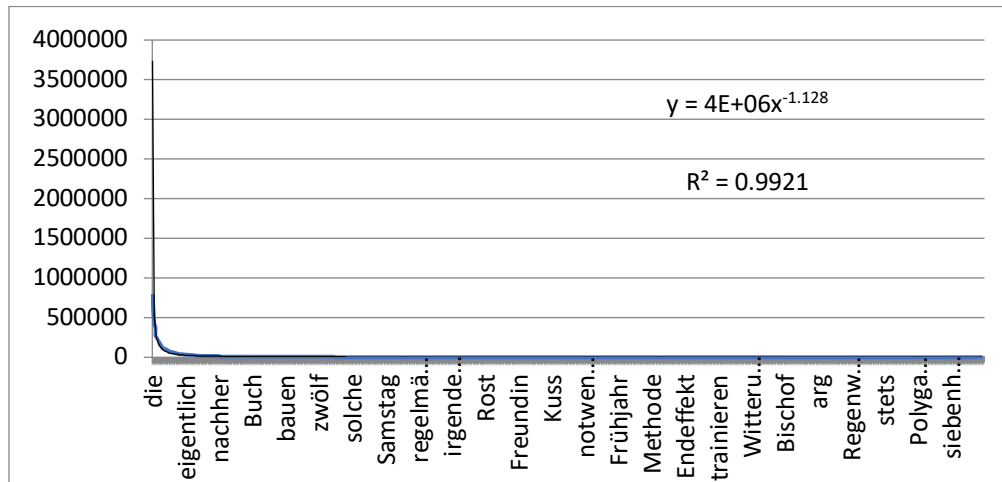
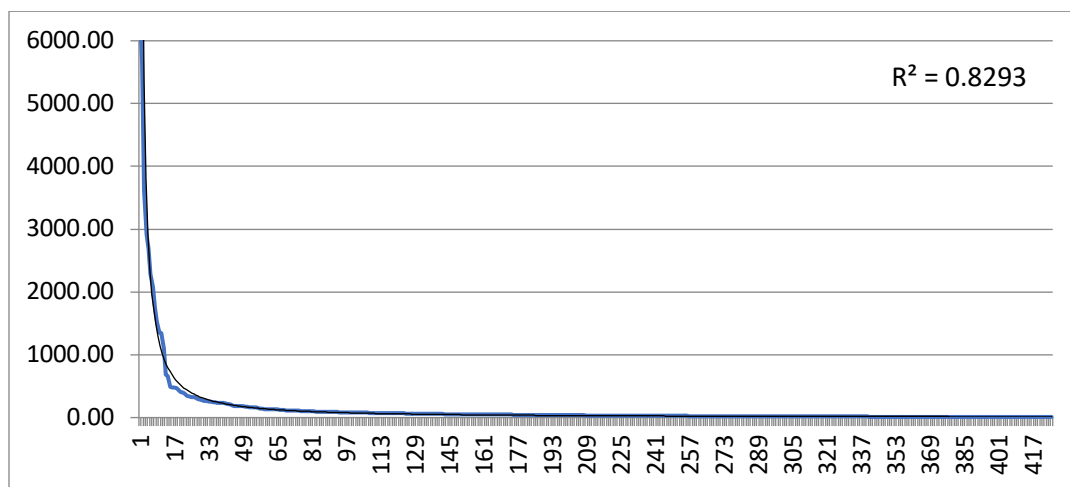


Figure 2. Rank order plot of vocabulary word frequency from the most frequent 425 students' vocabulary words. Frequency is shown on the y-axis, and the numerical identifier of the word is on the x-axis. The frequency range is from 6,486.01 to 10.01 words per million. The x-axis shows the count number that represents each word, from 1 to 425.



4.3.5 Why use corpora?

Examining corpora provides a method of evaluating the relative frequency of the words that students reported as salient. While it could be explained that students acquire relatively common words easily due to frequent exposure, the less common and infrequent words were a mystery as to why students found them to be salient. Corpora searches show the linguistic environment that these vocabulary words are used.

Table 2. A sample portion of the modified coding sheet used to compare student vocabulary words and their occurrences in the Falko, DGD spoken, and Sketch Engine corpora of both written and spoken German. The far-left column lists the students' vocabulary words and coding numbers. The next column to the right headed Textbooks contains the total vocabulary word count in the texts. Three columns headed by Falko, DGD spoken, and Sketch Engine contain the frequencies of the students' vocabulary words found in each corpus.

kind	▼	Textbooks	Falko	DGD spoken	Sketch Engine/per million		class	
0301bleiben		9	486	3127	178.99	180473	7	remain
0798jede		9	301	1643	176.78	90319	8	each
0055anders		9	462	8188	174.38	119664	7	different
0220bekannt		9	112	1889	166.62	170115	7	known,famous
1084schließlich		7	74	1570	164.71	82901	8	finally,eventually
1247überhaupt		9	343	8178	163.92	108173	8	at all
1230Trotz		6	98	382	158.50	73254	8	defiance
1141sonst		8	89	10313	153.56	82664	8	otherwise
0321damals		9	120	7161	141.39	128906	7	backthen
1042Richtung		8	74	3134	137.58	133291	7	direction
0681Gesellschaft		8	3026	851	136.50	109574	8	society
0003Abend		9	0	5179	134.00	124946	7	evening
1195tatsächlich		7	117	1538	131.74	91438	8	actually
1148Spaß		9	189	4601	129.78	62415	8	fun
0898Meinung		9	1564	1722	128.60	75647	8	opinion
1016Raum		6	42	1170	128.07	84083	8	spaceorroom
0355dürfen		0	129	1588	127.81	122535	7	allowed
0800jemand		8	304	3174	122.50	67386	8	someone
0801Junge		9	142	5041	120.79	36008	9	boy
0884manchmal		9	402	9201	119.17	50653	9	sometimes
0184Bedeutung		8	458	703	116.17	58097	8	meaning
1567zusätzlich		6	13	237	114.70	39393	9	additional
0923möchten		9	335	683	114.52	38461	9	wouldlike
0477Ergebnis		8	84	1137	113.55	101982	8	result
1131Sicherheit		8	94	429	112.85	93023	8	security
1580Zweite		8	9	2179	112.62	11454	11	second

5 Conclusion

Second language vocabulary acquisition can be explained by an analysis that employs both Cognitive-Functional insights into language use and mathematical insights into vocabulary selection. The data shows that students are not acquiring words according to a grammar-list model of vocabulary learning, but it is theorized that they are adding words into a mental network that can be measured in terms of a power-law, and power laws provide the mathematical descriptions of lexical neural networks.

This observation implies that learners are sensitive to all new vocabulary words, depending on the word's perceived consecutiveness to existing memory networks. This observation is in agreement with the functional linguistic explanations of the frequency effects in grammar.

5.1 Development of a theory for vocabulary acquisition

Categorizing words by their relative frequency as found in each of the corpora of spoken and written German provides a perspective on why the words were selected by students as salient. An analysis suggests that certain words are salient because they are new to the student, but they also fit into a framework as defined by Cognitive linguistics theory.

If the cognitive linguistic perspective is taken, then the total lexical vocabulary is considered to be a vast network with linkages to meanings and words, instead of being a long list that is accessed by a grammatical rules routine, and consequently students are noticing words that are missing from their currently developed memory network.

Explaining the events in a cognitive linguistic approach, it would be said that students are adding to their network of knowledge and use, and new words

that are supported by expressions that are already known. Students acquire vocabulary words in a predictable manner, which is consistent with the observations of Ellis (2002), Larsen-Freeman and Tedics (2016), Larsen-Freeman (2017), and Fillpovic and Hawkins (2019), and this vocabulary acquisition is power-law related. The interacting elements of the vocabulary system are a complex adaptive system (Ellis, 2002; Larsen-Freeman, 1997; MacWinney, 1999).

A functionalist-cognitive approach would say that students learn words that are presented in environments that have surrounding meaningful words, and those students know the lexical meaning and the semantic values of the supporting environment. Otherwise, the new word would not have meaning and would not have been noticed as salient.

5.2 Proposed theory of a vocabulary network

A concept that relates to vocabulary acquisition is that a system, such as language, produces a self-organizing criticality (Waldrop, 1992). This concept parallels closely with the functional usage-based views of grammar and the lexicon (Kövecses & Szabco, 1996; Larsen-Freeman, 1997; MacWhinney, 1999; Seidenberg & MacDonald 1999; Bybee & Hopper 2001; Christiansen & Chater, 2001; de Bot, Lowie, & Verspoor, 2007; Ellis 2007, 2008; Ellis & Larsen Freeman, 2009; Ellis, 2008). Power-law behavior explains scaling, the ability to grow and change (Mitchell, 2009), which in this research would be the ability of the mental vocabulary to increase. As part of this self-organizing process, mental nodes are created, which are connections or linkages to other words.

The vocabulary is a vast network of words and meanings, connected by links that are reinforced by use and by their connection to other words that are used. It is a complicated system of word frequency and word meaning in which word-use strengthens the connections. The word and its associated meanings

and collocations are a viable part of the working memory (Ellis & Wulff, 2015, 2019; Javadi, & Kazemirad, 2020; Cilliers, 2001; Larsen-Freeman, 2020; Beckner, Blythe, Bybee, Christiansen, Croft, Ellis, Holland, Ke, Larsen-Freeman, & Schoenemann, 2009; Steyvers & Tenenbaum, 2005; Meara, 2007; Wang, Deyne, & McKague, 2022; Conklin & Schmitt, 2012; Perek, 2023; Wilks & Meara, 2002, 2007; Wilks, Meara, & Wolter, 2005; Wilks, 2009; Hulstijn, 2020; Freeborn, Andringa, Lunansky, & Rispens, 2023, Lowie, Verspoor, & Seton, 2010).

What was thought to be a collection of vocabulary words has been shown to be the result of a systematic process of vocabulary addition. Each added vocabulary word was related to and connected to words that co-occurred in the sentence or utterance, otherwise it would not make sense, and would be rejected. Therefore, the likelihood that a word is assimilated is related to the salience of the words associated with it.

Students have a general knowledge of word frequency and listen for words that they hear in the context of words that they already know. New vocabulary words are being preferentially added as a function of the power-law model, and the student is not aware of this process. It is theorized that new vocabulary will be added to the students' lexicons based on how students recognize the new words as salient. Students noticed high frequency words, and this indicates that they became aware of new situations in which these words can be used, and so these words were added to the lexicon in a representation that included this new meaning. The medium frequency words that were recorded indicate that the students are adding words to the mental network based on the frequency of the words in the network. A medium frequency word is a candidate for several networks. Low frequency words were not learned in isolation; they had to be added to a network that relates the new word to network meanings and uses.

5.2.1 Illustration of the vocabulary word *dunkle* (dark)

The diagram below illustrates the mental network for the word graph for *dunkle*, an adjective meaning dark. This diagram was constructed using data from the *Universität Leipzig Wortschatz* corpora of 46,843,422 sentences and is consistent with the Cognitive Linguistics visualization of schema and frameworks.

Each word in the plot represents a node, which is the junction of word connections. (The property of nodes is predicted by power-law equations.) Each mental connection to a word is a node, and the more the connecting words, the stronger the connection and the more susceptible a word is to be added as new vocabulary to the network.

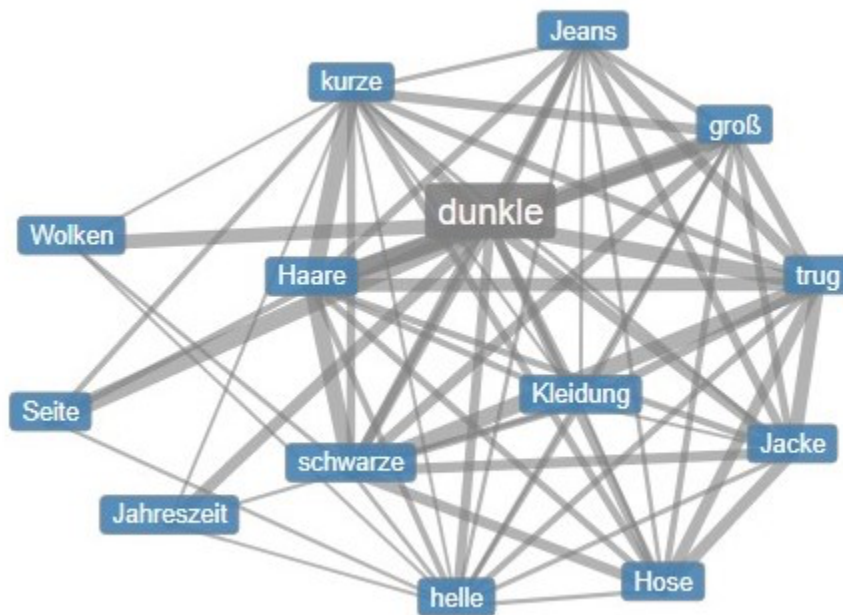


Figure 3. Vocabulary connections for the German word *dunkle*.

To illustrate the how words are added to the metal vocabulary, the primary word *dunkle* (dark) is shown by its connections (grey lines) to other words including *Kleidung* (clothing), *Haare* (hair), *Wolken* (clouds), *schwarze*

(black), *Jacke* (jacket), *Seite* (side), *Jeans* (jeans) and *Hose* (pants). These 13 words are not all the words associated with *dunkle*, but they are the most common words that are thought of when the word *dunkle* is mentioned, or as Functional linguists would say, when *dunkle* is activated. The 13 most related words are connected by lines of varying thickness and lengths indicating the relative frequency of use when the words are used together. This figure is also 3-dimensional and can be rotated in any direction, but the connections remain the same.

When a L2 learner hears the word *dunkle* in a sentence, for example *dunkle Wolken* (dark clouds), the mind activates it and all the words that the person knows that are associated with *dunkle*. If the specific use of *dunkle* is already known, then the vocabulary memory is reinforced.

If the L2 learner was exposed to *dunkle* as a new word in a sentence containing one or more of the 13 connected words, knowledge of the associated words would also speed activation and facilitate learning. The more likely the words in the neighborhood are known, the more likely the new word is to be learned and added to the network. Students do not learn vocabulary in isolation.

5.2.2 Summary: Two theories of vocabulary learning are proposed

The first theory is that L2 learners acquire new vocabulary consistent with a power-law explanation of the data. Learners acquire both very common words, moderately infrequent words, and rarely used words, according to a power-law distribution of the rank order of the word frequencies. Consistent with and part of power-law analysis is the understanding that the lexicon is power-law distributed and can also be depicted as a series of nodes, such as the one that is illustrated in the *dunkle* network in figure 3.

The second theory is based on the examination of the data is that students acquire vocabulary words on the edge of their expanding mental lexicon. That is to say, in order to acquire a new vocabulary word, the student must have in his or her mental lexicon related words and associated meanings that are used with the new word. Students acquire new words, not as a long vocabulary list, but as meaningful connections in a network. Therefore, learning is facilitated when new vocabulary words are used with the second language words that the student knows.

5.3 Future directions

This study was conducted with L2 German language students, but could be repeated with any second language program, or any first language program to determine if the neural network method of vocabulary learning is universal across language learning situations. Insights in how students learn vocabulary would motivate changes in textbook design.

If this study is repeated, the student's existing vocabulary networks should be assessed as a first step toward improving the theories offered.

Wang and Christiansen (2024) offer insights into teaching word chunks in the second language, called collocations in section 4.2.1. The Unified Competition Model they discuss might be reconciled with this research data.

References

- Achard, M. (2018). Teaching usage and concepts: Toward a cognitive pedagogical grammar 37-62 in Andrea Tyler, Lihong Huang, Hana Jan (Eds.) *What is Applied Cognitive Linguistics? : Answers from Current SLA Research*, edited by Andrea Tyler, et al., De Gruyter, Inc.
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W. & Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, 59, 1-26.
- Boers, F. (2013). Cognitive Linguistic approaches to teaching vocabulary: Assessment and integration. *Language Teaching*, 46.2. 208-224.
- Boers, F. & Lindstromberg, S. (2008). *Cognitive Linguistic Approaches to Teaching Vocabulary and Phraseology*. Berlin: Mouton De Gruyter.
- Boers, F. & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics* 32, 83-111.
- Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.
- Capocci, A., Servedio, V.D. P., Colaioni, F., Buriol, L. S., Donato D., Leonardi, S. & Galdarelli, G. (2006). Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia. *Physical Review E* 74, 036116.
- Chamot, A. U. (2005). Language learning strategy instruction: Current issues and research. *Annual review of applied linguistics*, 25, 112-130. Christiansen, M. & Chater, N. (Eds.). 2001. *Connectionist psycholinguistics: Capturing the empirical data*. *Trends in Cognitive Sciences*, 5(2), 82-88.
- Chun Tie, Y., Birks, M., & Francis, K. (2019). Grounded theory research: A design framework for novice researchers. *SAGE open medicine*, 7, 1-8.

- Chun, Y., Birk, M., & Francis, K. (2019) Grounded theory research: A design framework for novice researchers. *Sage Open Medicine*. Vol 7: 1-8.
- Cilliers, P. (2001). Boundaries, hierarchies and networks in complex systems. *International Journal of Innovation Management*, 5(02), 135-147.
- Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4), 661-703.
- Common European Framework for Reference for Languages, The. (2022). <https://www.examenglish.com/CEFR/cefr.php>, (2022).
- Conrad, C. (1982). Grounded Theory: An alternative approach to research in higher education. *The Review of Higher Education*. 5(4), pp 239-249.
- Conklin, C., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than non-formulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72-89.
- DeKeyser, R. (1991). Foreign language development during a semester abroad. In *Foreign Language Acquisition: Research and the Classroom*, F. F. Freed (Ed.), (pp.104-119). Lexington, MA: D. C. Heath.
- de Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory to Second Language Acquisition. *Bilingualism: Language and Cognition*, 10: 7-21.
- Eigen, M., & Winkler, R. (1981). *Laws of the Game. How the Principles of Nature Govern Chance*. Alfred A. Knopf, NY.
- Ellis, N. & Sagarra, N. (2011). Learned attention in adult language acquisition. A replication and generalization study and meta-analysis. *Studies in Second Language Acquisition* 33, pp. 589-624.
- Ellis, N. (2006). Cognitive perspectives on SLA: The associative-cognitive CREED. *AILA Review* 19, pp.100-121.
- Ellis, N. (2002). Frequency Effects in Language Processing. *SLA* 24, pp. 143-188.

- Ellis, N. (2008). Usage-based and form-focused language acquisition. The associative learning of constructions, learned attention, and the limited L2 endstate. In P. Robinson & N. Ellis (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*. (pp. 372-405). New York: Routledge.
- Ellis, N., & Larsen-Freeman, D. (2009). Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage. *Language Learning*, 59, 90-125.
- Ellis, N & Wulff, S. (2019). Cognitive Approaches to Second Language Acquisition. 41 Schwieter, J. and A. Benati. (2019). *The Cambridge Handbook of Language Learning*. Cambridge University Press. Cambridge, UK.
- Filipović, L. & Hawkins, J. (2019). The complex adaptive system principles model for bilingualism: Language interactions within and across bilingual minds. *International Journal of Bilingualism*, 23(6), 1223-1248.
- Freed, B., S. So, & N. Lazar. (2003). Language learning abroad: How do gains in written fluency compare with oral fluency in French as a second language? *ADFL Bulletin* 34 (3): 34-40.
- Freeborn, L., S. Andringa, G. Lunansky & J. Rispen. (2023). Network analysis for modeling complex systems. *SLA research Studies in Second Language Acquisition* (2023), 45, 526–557 doi:10.1017/S0272263122000407
- Friedman, D. (2023). Qualitative and Interpretive Approaches to Data in L2 Research 51-71 In A. Mackey and S. Gass. *Current Approaches in Second Language Acquisition Research*. Wiley Blackwell, Hoboken NJ.
- Gass, S., Cohen, A. & Tarone E.. (1994). *Research Methodology in Second-Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Geeraerts, D. (2008). Introduction. A rough guide to Cognitive Linguistics. In D. Geeraerts (Ed.), *Cognitive Linguistics: Basic Readings*. Berlin: Mouton de Gruyter, pp.1-28.

- Glaser, B. (1978). *Theoretical sensitivity*. Mill Valley, CA: The Sociology Press.
- Glaser, B. and Strauss, A. (1967). *Discovery of grounded theory: strategies for qualitative research*. New York: Aldine de Gruyter.
- Hashemian, M., & M. Nehzad. (2007). The development of conceptual fluency & metaphorical competence in L2 learners. *Linguistik Online*, 30, 41-56.
- Haukås, Å., Bjørke, C. & Dypedahl, M. (2018). *Metacognition in language learning and teaching*. Taylor & Francis.
- Hulstijn, J. (2005). *Theoretical and empirical issues in the study of implicit and explicit second-language learning*. Cambridge University Press 0272-2631005
- Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a foreign language*, 17(1), 23-59.
- Hutchinson, S. (2005). Education and Grounded Theory. In Robert Sherman and Rodman Webb. *Qualitative Research in Education: Focus and Methods*. Routledge Falmer: New York.
- Javadi, Y., & Kazemirad, F. (2020). Usage-based approaches to second language acquisition: Cognitive and social aspects. *Journal of Language Teaching and Research*, 11(3), 473-479.
- Keating, Gregory. (2021). Online methods in research on input processing and processing instruction. In *Research on Second Language Processing and Processing Instruction: Studies in Honor of Bill VanPatten*, Michael J. Leeser, et al., eds. John Benjamins Publishing Company.
- Kohl, K. (2007). *Poetologische Metaphern*. Walter de Gruyter, Berlin.
- Kovecses, Z., & Szabco, P. (1996). Idioms: A view from cognitive semantics. *Applied Linguistics*, 17(3), 326-355.
- Kumar, A. A., Steyvers, M., & Balota, D. A. (2022). A critical review of network-based and distributional approaches to semantic memory structure and processes. *Topics in Cognitive Science*, 14(1), 54-77.

- Laufer, B. (2001). Reading, word-focused activities and incidentally vocabulary acquisition in a second language. *Prospect*, 16(3), 44-54.
- Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review*, 59(4), 567-587.
- Lee, B. (2021). The Duck-Rabbit Problem and Linguistic Questions. *Sixth Annual Law and Corpus Linguistics Conference*, Feb 5, 2021. Provo, UT.
- Li, J., & Wang, C. (2024). Mediating roles of motivational beliefs and vocabulary learning strategies for the relationship between self-regulation and vocabulary proficiency. *European Journal of Education*, 59(4), e12706.
- Lightbown, P. (2000). Anniversary article. Classroom SLA research and second language teaching. *Applied linguistics*, 21(4), 431-462.
- Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied linguistics*, 18(2), 141-165.
- Larsen-Freeman, D. (2020). Complex dynamic systems theory. In *Theories in second language acquisition* (pp. 248-270). Routledge.
- Larsen-Freeman, D. (2023). Complex dynamic systems theory: A webinar with Diane Larsen-Freeman. *Language Teaching*, 56(3), 402-419.
- Li, Q. (2024). Cognitive Linguistics and vocabulary Acquisition: Enhancing English Learners' Mental Lexicon. *Education Insights* 1(2):55-63. DOI:10.70088/9qmx7977.
- Loewen, S., Tuzcu, A. & Philp, J. (2023). Classroom Research Methods 11-32. In A. Mackey and S. Gass. *Current Approaches in Second Language Acquisition Research*. Wiley Blackwell, Hoboken NJ.
- Lowie, W., Verspoor, M., & Seton, B. (2010). Conceptual representations in the multilingual mind. *Converging Evidence in Language and Communication Research (CELCR)*, 135.

- MacWhinney, B. (Ed.). (1999). *The Emergence of Language*. Hillsdale, NJ: Erlbaum.
- Madlener, K. (2015). *Frequency Effects in Instructed Second Language Acquisition*. De Gruyter Mouton: Berlin.
- Maxwell, J. (2013). *Qualitative Research Design. An Interactive Approach*. Los Angeles: Sage.
- Meara, P. (2005). Designing vocabulary tests for English, Spanish and other languages. In C. Butler, M. Gómez-González & S. Doval Suárez (Eds), *The dynamics of language use* (pp. 271-286). Amsterdam: Benjamins.
- Meara, P. (2007). Simulating word associations in an L2: The effects of structural complexity. In *Language Forum* (Vol. 33, No. 2, pp. 13-31).
- Meara, P. (2009). *Connected words: Word associations and second language vocabulary acquisition*. John Benjamins Publishing Company.
- Wray, A. (2000).
- Miles, M. A. Michael Huberman, M. & Saldaña, J. (2014). *Qualitative Data Analysis*. Sage Publications: Thousand Oaks, CA.
- Mitchell, M. (2009). *Complexity. A guided Tour*. Oxford: Oxford University Press.
- Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston: Heinle.
- Nation, I.S.P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle.
- Nation, P. (2001, 2022). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, P. and Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31 (7), (pp. 9-13).
- Nattinger, I. & DeCarrico. J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

- Pellicer-Sánchez, A., & Boers, F. (2018). Pedagogical approaches to the teaching and learning of formulaic language. *Understanding formulaic language*, 153-173.
- Pérez-Vidal, C & Juan-Garau, M. (2009). The effect of study abroad (SA) on written performance. John Benjamins Publishing Company. *EUROSLA Yearbook 9*. 269-295.
- Richer, S. (1975). School effects: The case for grounded Theory. *Sociology of Education*, 48 (Fall), 383-399.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge, UK: Cambridge University Press.
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12, 329–363.
- Seidenberg, M. & MacDonald, M. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, 23(4), 569-588.
- Steen, G. (Ed.) (2018) *Visual Metaphor: Structure and Process*. Amsterdam: John Benjamins Publishing Company. Doi.org/10.1075/celcr.18
- 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.
- Taylor, J. R., & Littlemore, J. (2014). *The Bloomsbury companion to cognitive linguistics*.
- Tie, Y. C., Birks, M. & Francis, K. (2019) *Grounded theory research: A design framework for novice researchers*. Sage Open Medicine. Vol 7: 1-8.
- Tomasello, M. (2003). *Constructing a Language*. A Usage-based Theory of Language Acquisition. Cambridge: Harvard University Press.
- Waldrop, M. (1992). *Complexity*. Simon and Schuster. New York.

- Wang, S. Y., & Christiansen, M. H. (2024). Chunking in the second language: Implications for language learning and teaching. *Language Teaching Research Quarterly*, 44, 84-106. <https://doi.org/10.32038/ltrq.2024.44.09>
- Wang, A., De Deyne, S., McKague, M., & Perfors, A. (2022). Core words in semantic representation. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 44, No. 44).
- Webb, S., Uchihara, T., & Yanagisawa, A. (2023). How effective is second language incidental vocabulary learning? A meta-analysis. *Language Teaching*, 56(2), 161-180.
- Webb, S., & Chang, C-S. A. (2012). Second language vocabulary growth. *RELC Journal*, 43, 113–126.
- Wenden, A. L. (1998). Metacognitive knowledge and language learning¹. *Applied linguistics*, 19(4), 515-537.
- West, G. B., Brown, J. H., & Enquist, B. J. (1999). The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science*, 284(5420), 1677-1679.
- West, G. B., Brown, J. H., & Enquist, B. J. (2000). The origin of universal scaling laws in biology. *Scaling in biology*, 87-112.
- Wilks, C., Meara, P., & Wolter, B. (2005). A further note on simulating word association behaviour in a second language. *Second Language Research*, 21(4), 359-372.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press, 3-93.
- Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied linguistics*, 21(4), 463-489.

Appendix 1: Corpora used in the data evaluation

Database for Spoken German (Datenbank für Gesprochenes Deutsch, DGD). The DGD corpus is available from: <https://dgd.ids-mannheim.de/dgd/pragdb.dgd>.

The Database for Spoken German (DGD).is a corpus platform of the Archive for Spoken German at the Leibniz Institute for the German Language. The *Datenbank für Gesprochenes Deutsch* corpus portions that were used contain slightly more than 12 million words.

The Falko Corpus is found online at <https://hu-berlin.de/falko>. <https://korpling.german.hu-berlin.de/falko-suche/> Annis interface was used.

The Falko Corpus is an error-annotated learner corpus of German as a foreign language. Falko is sponsored by *Humboldt-Universität zu Berlin, Institut für deutsche Sprache und Linguistik*. For purposes of this research, the corpus was composed of 12 applicable sub-corpora totaling 1,246,087 words.

The Leipzig Corpora Collection (Wortschatz Leipzig) is available from <https://wortschatz.uni-leipzig.de/en/download>.

The Leipzig Corpora Collection (*Wortschatz Leipzig*) was used to evaluate the co-occurrences of the students' vocabulary words. Co-occurring words are those that often occur with the target vocabulary word and are in the same sentence. These corpora employ the log-likelihood ratio as a significance measure (Goldhahn, Eckert & Quasthoff, 2012). Corpora collections contain randomly selected sentences in sizes from 10,000 sentences to 1 million sentences. The sources are either newspaper texts or texts randomly collected from the web.

The Sketch Engine is available from:

Lexical Computing Limited, 2003: <https://www.sketchengine.eu>.

The Sketch Engine (Lexical Computing Limited, 2003) is a corpus analysis tool that is designed for text analysis and text mining applications. The 16,526,335,416-word German Web 2013 (deTenTen13) sub-corpus, based on web crawling, was selected for this research project.

Appendix 2: Textbooks used in the summer language program

Buscha, Anne and Szilvia Szita. (2013). Begegnungen, B1. Leipzig. Schubert-Verlag. (72,891 words),

Buscha, Anne and Szilvia Szita. Begegnungen B1 + Lösungsschlüssel. (answer booklet). (19,738 words)

Buscha, Anne and Susanne Raven. (2010). Erkundungeng, B2. Leipzig. Schubert-Verlag. (114,914 words)

Buscha, Anne and Sisamme Ravem. Begegnungen B2. Lösungsschlüssel. (87,280 words)

Buscha, Anne, Susanne Raven, Szilvia Szita. (2016). Erkundungen, C1. Leipzig. Schubert-Verlag. (179,613 words),

Buscha, Anne, Susanne Raven, Szilvia Szita. Erkundungen, C1, Lösungsschlüssel. (133,808 words)

Buscha, Anne, Susanne Raven, Mathias Toscher. (2014). Erkundungen, C2. Leipzig. Schubert-Verlag. (193,568 words)

Buscha, Anne, Susanne Raven, Mathias Toscher. Erkundungen, C2, Lösungsschlüssel. (79,004 words).