Corpus-driven vocabulary learning! – On the design, implementation and evaluation of persuasive tools for computer-assisted vocabulary learning using Role and Reference Grammar

By

Judith Gottschalk, B.A.

Submitted in Fulfillment of the Requirements for the M.A. program in Human-Centered Informatics at the Department of Communication and Psychology at Aalborg University Denmark

Nuremberg, Germany
December 21th 2014

Supervisor: Prof. Dr. Peter Øhrstrøm
Co-Supervisor: Dr. Ulrik Sandborg-Petersen
Characters: 172.745 (71,87 pages)

Signature:

Judith Gottschalk

Meinem Vater Volker Gottschalk in
liebevoller Erinnerung gewidmet

Title: Corpus-driven vocabulary learning! - On the design, implementation and evaluation of persuasive tools for computer-assisted vocabulary learning using Role and Reference Grammar

Author: Judith Gottschalk (study no. 2012245)

Date: December 21th 2014

Study: Human-Centered Informatics

Term: 10th

Supervisor: Prof. Dr. Peter Øhrstrøm

Co-Supervisor: Dr. Ulrik Sandborg-Petersen

Characters: 172.745 (71,87 pages); summary, table of contexts, references and appendixes not included The physical page count is higher due to the layout.

Summary

Winther-Nielsen (ms) lists vocabulary acquisition as one aspect of the development of Bible Online Learner which can be improved. The problem I address in this master's thesis is the following: When learning Biblical Hebrew, what causes most learners serious problems is remembering and correctly applying the huge amounts of vocabulary. Many lexemes have several different meanings and, when translating from the Hebrew Bible, it is nearly impossible for the students to choose the meaning which correctly fits in the context. How can this problem be solved with Bible Online Learner and Learning Journey?

Learning Journey Online is, like Bible Online Learner, a web application written in PHP and JavaScript, and it can measure the learning progress of students by using Item Response Theory [IRT] within the persuasive architecture of Bible Online Learner. With IRT it is possible to statistically measure the performance of a learner in order to predict the probability of her response by establishing the position of the individual learner along a line of some latent dimension. The use of IRT is quite prominent in educational environments and in this context the latent trait is often called *ability*. Moreover, it is also possible to measure how difficult it is to acquire an item which is to be learned with IRT (cf. Parchev 2004: 5).

This master's thesis is organized in an unusual way: It consists of a general introduction dealing with a case study on teaching Biblical Hebrew in a class taught by Nicolai Winther-Nielsen at Fjellhaug International University College in Copenhagen and it gives an overview about Biblical Hebrew didactics and learning strategies for learning Biblical Hebrew vocabulary. The heart of this thesis is however formed by two publishable papers on the theoretical linguistics basis for my approach to improved vocabulary learning with Bible Online Learner and Learning Journey and one on my concrete design approach which uses the theoretical foundation developed in the first paper.

In the course of my master's studies I have participated the International Role and Reference Grammar Conference 2013 in Freiburg, Germany, and presented a paper on Learning Journey and methods to automatically derive the meaning of verbs from the corpus used in Bible Online Learner using conceptual graphs as an alternative to the classical Aktionsarten used by Van Valin (2005) in RRG. The paper is accepted for publication in the in NIHIN, an open access series of Freiburg University dealing with linguistics. My paper for the RRG

proceedings is the theoretical foundation to my approach to computer-supported vocabulary learning. It deals with the question of how computer-supported vocabulary learning can be improved by using a computationally adequate model of RRG. It suggests conceptual graphs (cf. Sowa 2000) as a new approach to semantics in RRG to develop a computationally tractable version of RRG which can be implemented in a semantic parser as an extension of existing learning software for Biblical Hebrew. The application of conceptual graphs has the advantage that computational approaches for ancient languages such as Biblical Hebrew can be developed. In this approach, a linking algorithm from syntax to semantics is reduced to a set of lexical rules which match attribute value matrices defining the layered structure of the clause against an ontology (cf. Gottschalk 2010; Gottschalk 2012a; Gottschalk in press), and uses an algorithm for the automatic determination of Aktionsarten for Biblical Hebrew developed by Winther-Nielsen (ms). With this semantic parser it is possible to derive glossing for Bible Online Learner which use exact meaning of a Hebrew word depending on their morphosyntactic context. The current problem with the glossing in the corpus Bible Online Learner is using, is namely that several Hebrew words have more than one meaning and that their means are all displayed in the glossing in Bible Online Learner.

This way it is difficult for the student to learn Hebrew vocabulary by reading the Bible with Bible Online Learner as they learners cannot easily derive lexical rules which help them to determine a proper translation of a lexeme which has multiple meaning depending on its context. My hypothesis 1) regarding computer-supported vocabulary learning is that words are best learned within the morphosyntactic context they occur and to motivate the learners to explore the hypertexed corpus of the Hebrew Bible by themselves. The idea regarding an improved glossing in Bible Online Learner is: If the learners are only exposed to ideal glossings and to perfect examples they will be able to derive lexical rules for the determination of the meaning of words which have multiple meaning depending on the morphosyntactic context they occur in. This is what I discuss in the second publishable paper, which is part of this thesis: I am dealing with the question of how the learning with Bible Online Learner and Learning Journey can be improved towards computer supported vocabulary learning and how the theoretical findings from the first paper, Gottschalk (accepted), which is part of this master's thesis, and Winther-Nielsen (ms) can be practically applied within the tools to improve facilitation and learning? The second paper presents a design on a practical approach to how learners using Bible Online Learner will be able to acquire lexical rules to determine the exact meaning of *nātan* in the corresponding

morphosyntactic context. It will one the hand show how exercises in Bible Online Learner needs to be designed in order to enable corpus-driven language learning and on the other hand I will show how the sound learning statistics in Learning Journey can be used for vocabulary learning. Here my hypothesis is the following: While tools like Anki and Memrise use specific-fixed algorithms which determine when recently acquired vocabulary needs to be repeated in order to be learning (cf. Edge et al. 2012), Learning Journey with its statistical framework can offer a much more flexible approach which is user-driven and enables tailoring and tunneling, which are two persuasive means, which following Gottschalk and Winther-Nielsen (2013) should enable successful language learning with Bible Online Learner. In this paper I am going to discuss the algorithms used in Anki and other tools and compare them with the approach chosen in Learning Journey. More important however is my idea that learning the vocabulary of Biblical Hebrew is best possible within contextualized vocabulary learning as done with Bible Online Learner. I present a study made with Nicolai Winther-Nielsen's class in Copenhagen and am able to show that those students who learned the vocabulary of Biblical Hebrew contextualized with Bible Online Learner were more successful in learning the language than those students who learned vocabulary decontextualized with Memrise.

With these two papers I am on the one hand developing a formally solid framework for semantic parsing and automatic meaning determination for a specific kind of verbs within a linguistic corpus and doing work in the area of theoretical functional computational linguistics and on the other hand I am developing a persuasive approach to computer-supported vocabulary acquisition within the area of applied computational linguistics, which practically enables learners to improve their language skills.

Acknowledgements

I am deeply indebted to my father, Volker Gottschalk, who has supported me in so many ways during his lifetime. Even now while being in heaven he is taking care of my mother, my two brothers and me. Thanks for all you have done for me. I am grateful to my mother, Elke, and my brothers Maximilian and Claudius for their support and encouragement. Thank you all so much.

I would like to express my greatest gratitude to my supervisors (in alphabetic order) Prof. Dr. Peter Øhrstrøm, Dr. Ulrik Sandborg-Petersen, Claus Tøndering and Prof. Dr. Nicolai Winther-Nielsen for many insightful conversations during the development of the ideas in this thesis, and for helpful comments on the text. I thank Prof. Dr. Peter Øhrstrøm for his reliable guidance during my time as distance student at Aalborg University and for his encouragement with the writing of this thesis. Prof. Winther-Nielsen has supported me not only in writing this thesis but is also a wonderful friend to me who gave me the emotionally strength on my way to finishing my master's degree. He was there for me in the most difficult time in 2014 when my father died and encouraged me to not give up. Specifically I would like to thank my programming supervisor Claus Tøndering who supporting me in developing the important programming skills which led me to the implementation of Learning Journey and which are also of greatest importance for my future career as a professional programmer. I thank you for your patience and willingness to discuss all kinds of programming questions with me and for being a wonderful and encouraging teacher.

Further I would like to say thank you to my friend and mentor Dr. Brian Nolan for inspiring me with new ideas when it comes to the question of the computability of Role and Reference Grammar and for his support while working on the RRG part of this thesis. Thank you for your friendship and also for the emotional support you gave me during the past six years since we met the first time in Dublin. I express me thanks to Christian Højgaard for sharing his teaching experiences at SALT in Madagascar with me and to my friend Dr. habil. Hagen Langer with whom I discussed questions of the computability of RRG for several years. I would also like to say thank you to Prof. Dr. Robert D. Van Valin, Jr. for answering specific questions on RRG.

Thanks to Rebecca Jackson for her professional proofreading of this master's thesis and for giving me valuable tips and hints to improve my writing. You did a marvelous job!

Finally thanks to all my friends who supported me during my studies: Martin Braml, Carina Heßeling, Dr. Ulrike Hinke-Dörnemann, Regina Froschauer, Kim Hülsewede, Sabrina Kirschner, Manuel Landinger, Matthias Ströll, Micoto Szillat, Brigitte Thaler, Ute Treschan, Uwe Schade and Andreas Sommer. I would like to express my gratitude to all those who gave me the encouragement to complete this thesis.

Table of Contents

Computer-supported vocabulary learning with Bible Online Learner and Learning Journey – The introduction

Judith Gottschalk
Aalborg University, Rendsburggade 14, 9000 Aalborg
gottschalk.judith@gmail.com

1 Introduction

"I cannot teach you anything, the Bible will be your tutor". This is how Nicolai Winther-Nielsen opened the first lecture in his introductory Hebrew course at Fjellhaug International University College Denmark [FIUC]. Seventeen students of theology, female and male, aged between 18 and 80, are sitting in his class in Copenhagen while he explains how he plans to teach them Biblical Hebrew within the coming months: Winther-Nielsen is using Bible Online Learner and Learning Journey, two persuasive tools for computer supported language learning.

Over the past three years, six partner institutions from four countries in Europe have been working on the development of Persuasive Learning Objects and Technologies (PLOT) in a Lifelong learning Project, which is called EuroPLOT (www.eplot.eu). The Education, Audiovisual and Culture Executive Agency (EACEA) of the European Commission funded EuroPLOT from 2010 till 2013. Within this project, PLOTLearner has been developed as tool for persuasive language Learning.

PLOTLearner has been repurposed as an online application developed by Claus Tøndering under the name of Bible Learner Online (http://bibleol.3bmoodle.dk/; Winther-Nielsen 2013a: 22f). With Bible Online Learner, his students can, on the one hand, explore a hypertext corpus of the Hebrew Bible to improve their knowledge of Biblical Hebrew via reading. Furthermore, on the other hand they can complete various grammar exercises to improve their language skills with practical drills.

Additionally, Winther-Nielsen as facilitator also uses a second tool for his language teaching:

Learning Journey. Learning Journey is, like Bible Online Learner, a web application written in PHP and JavaScript, and it can measure the learning progress of students by using Item Response Theory [IRT] within the persuasive architecture of Bible Online Learner. With IRT it is possible to statistically measure the performance of a learner in order to predict the probability of her response by establishing the position of the individual learner along a line of some latent dimension. The use of IRT is quite prominent in educational environments and in this context the latent trait is often called *ability*. Moreover, it is also possible to measure how difficult it is to acquire an item which is to be learned with IRT (cf. Parchev 2004: 5).The guiding hypothesis, which Learning Journey Online is based on, is that gathering data from learning statistics through computational surveillance and data mining algorithms and their visualization can support persuasive teaching in systems for computer-assisted language learning.

When students begin their Hebrew lessons with Nicolai Winther-Nielsen at the start of each semester, they realize they are learning a language which is no longer spoken, and which differs typologically from Danish, English and German, for example. This leads students to raise the following question with Professor Winther-Nielsen: “Why do we need to learn Hebrew today?”

In many departments of theology and seminaries all over the world, the consensus concerning why the ancient language of Biblical Hebrew is still taught is that the knowledge of Biblical Hebrew should enable the students to do linguistically based exegesis. With a knowledge of Hebrew, learners can investigate the texts of the Old Testament from a linguistic perspective to unfold the meaning of a passage by employing a meaning-based analysis which is informed by the syntactic structure of the text. Another reason for adding Biblical Hebrew to the theology curriculum is that knowledge of Biblical Hebrew is indispensible for the translation of the Old Testament. One might be surprised that the Bible is still translated but even these days, there is a need for new translations for two reasons a) the language of Bible translations need to be constantly updated to reflect language change and b) there are still a considerable number of languages without a proper Bible translation. This means that learning an ancient language like Biblical Hebrew is still an issue in present theology.

And what does it mean to know Biblical Hebrew if it is not spoken anymore? While in modern languages like English or Danish production and comprehension are included in L2

acquisition, the objective of learning ancient languages like Biblical Hebrew is the ability to read text written in this language (cf. Skovenborg ms). Being able to read Biblical Hebrew requires knowledge of the scripture, as the language is not written in the Latin alphabet, and being able to read also requires the ability to parse Hebrew morphology and syntax.

The Hebrew alphabet has 23 consonants and, additionally, the writing system uses 13 vowels. Since Hebrew is rarely transliterated, it is absolutely vital to learn the Hebrew scripture. Hebrew morphology looks as follows: verbs, nouns or adjectives use consonantal roots of three (or sometimes two) letters. Affixes and vocalizations are added to form words. What is specifically challenging about Hebrew morphology is its verbal system: It consists of seven stems with 30 forms each (cf. Skovenborg ms).

In his introductory course, Winther-Nielsen regularly spends a considerable amount of time on teaching Hebrew morphology to the students. Here, the use of Bible Online Learner is very helpful for the students as, on the one hand, they can explore the corpus of the Hebrew Bible in Bible Online Learner to learn parsing via incidental reading and, on the other hand, they can practise Hebrew grammar with the tool.

Compared to the morphology, Hebrew syntax is relatively simple. Most Hebrew clauses are coordinated, not subordinated. However, this means expressing oneself as in an Indo-European language is not possible, which can be confusing for Western students (cf. Skovenborg ms).

Winther-Nielsen (ms) lists vocabulary acquisition as one aspect of the development of Bible Online Learner which can be improved. The problem I address in this master's thesis is the following: When learning Biblical Hebrew, what causes most learners serious problems is remembering and correctly applying the huge amounts of vocabulary. Many lexemes have several different meanings and, when translating from the Hebrew Bible, it is nearly impossible for the students to choose the meaning which correctly fits in the context. Students have various strategies when learning their vocabulary: Some write hundreds of record cards by hand with the Hebrew lexeme on the front and the translations in their native language on the back so that they can use them to recall the foreign words. Others use technical means. Computer programs like Anki (http://ankisrs.net), where one can generate digital record cards called flash cards, are commonly used, or a webpage called Memrise

(http://www.memrise.com) is consulted. Here the students can learn their vocabulary in a gameful way.

The roots of Bible Online Learner and Learning Journey lie back in 2010. This is when the EuroPLOT project first proposed persuasive technology for language learning and knew that quiz technology as found in a tool called 3ET can support learners with reading in a new language (cf. Winther-Nielsen 2013a; Winther-Nielsen 2011). While 3ET was a tool which motivated students to practice and complete exercises, it lacked a stronger persuasive capability to change users' behavior and motivations (cf. Winther-Nielsen 2013a). Following Winther-Nielsen, Bible Online Learner in turn is developed to allow students to:

- practice skills in reading, writing, and parsing for proficiency
- explore the structure of Biblical Hebrew for engagement in a simulation of the grammar of the text
- exploit resources activated by the text for a virtual world which functions as a learning environment

(cf. Winther-Nielsen 2013a)

Bible Online Learner is a persuasive tool and it motivates learners using conditioning through certificates, surveillance by teachers and peers, self-monitoring and learning recommendations. This way, the learners become autonomous language learners (cf. Winther-Nielsen 2013a). Winther-Nielsen characterizes the technology in different ways based on the dimensions of language drilling, text display, reading and vocabulary learning. It is:

- a tool that simplifies the practice involved to acquire morphology (drills)
- a virtual tutor simulating the study of grammar in a text
- an interlinear guide with translation (pronunciation support)
- a reading helper to check typing, reading and spelling skills (self-corrective reading)
- a translation assistant with glosses and with the ranking of word frequency (vocabulary training)
- a virtual world displaying material from the cultural background (interpretation)

(cf. Winther-Nielsen 2013a)

The first step in language learning with Bible Online Learner is using its capability for skills

training and its way of displaying the learning progress through statistics.

In Bible Online Learner, learners receive access to numerous quizzes containing exercises dealing with the various grammatical phenomena which a facilitator has discussed in class. These exercises are generated by a facilitator like Nicolai Winther-Nielsen. This way, language learning with Bible Online Learner is flexible and can be adjusted to different curricula. Figure 1 shows a screenshot of a quiz in Bible Online Learner dealing with parts of speech:

Figure 1 Quiz in Bible Online Learner on the correct choice of parts of speech in Biblical Hebrew

Here, a learner has to determine which part of speech the lexemes have in their grammatical context. By clicking on the right part of speech and solving the problems, the learners practice the grammatical phenomenon in question and submit data to a statistics server.

It is important to understand the role of practice, and it can best be understood within the pedagogical theory developed by Laurillard (2012) and applied by Winther-Nielsen (2013a). Laurillard's figure is reproduced with minor changes from Winther-Nielsen (2013a). In the

figure, the learner uses her personal goals and her current conceptual organization to select her current practice to generate actions in the external environment. In Bible Online Learner, the learner can use an action, in this case an exercise, which has been modeled by a teacher to build her practice capability (cf. Winther-Nielsen 2013a). The feedback the learner receives from either the facilitator or the learning environment modulates their concept, personal goals or their current practice capability. This way, new actions in the continual interactive process of learning are generated (cf. Winther-Nielsen 2013a; cf. Laurillard 2012). Laurillard (2012) explains how skill training with Bible Online Learner works as follows: "the shading identifies the internal cognitive components the learner deploys during the process, which the teacher is trying to influence" (Laurillard 2012: 61)." What the figure shows is that the learners' practice capability depends on an action modeled by the learning environment (B_E). In Bible Online Learner, learners interact with the external environment (E) through a web interface with the ETCBC database, containing the text in the target language (cf. Winther-Nielsen 2013a). The learning content is generated by the computational linguists maintaining the ETCBC database as well as the course facilitators through conceptual organization (C_T). This means that the learners' actions in achieving a knowledge of Biblical Hebrew (D_E) are embedded within a larger environment of facilitation (cf. Winther-Nielsen 2013a).

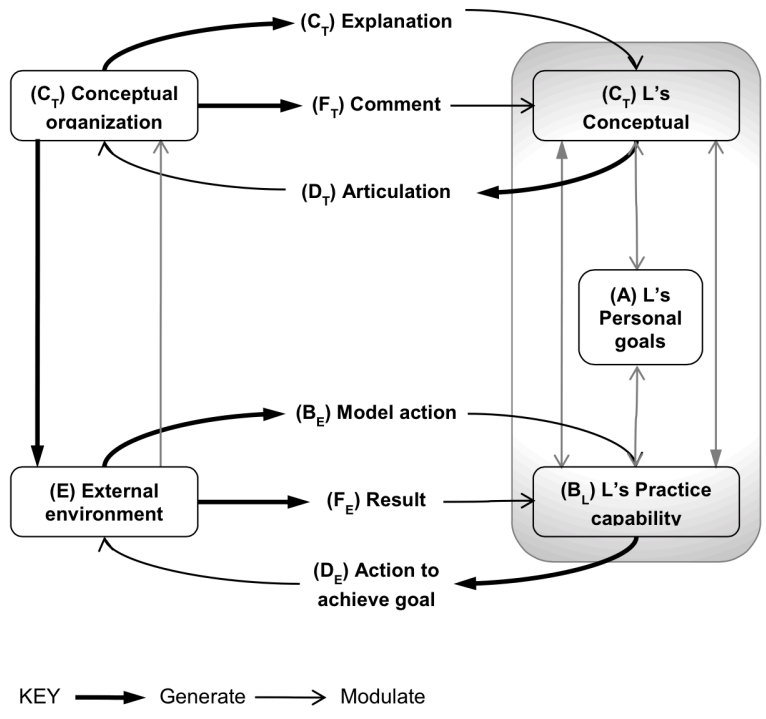

Figure 2: Learning circle, where the learner learns from an environment (B_E) and from

a teacher (Laurillard, Fig 4.1)

The crucial feature of Bible Online Learner is its ability to provide learners with immediate feedback. The result function (F_E) in Laurillard's model in figure 2 is realized by checking answers in the database and generating feedback. The learner can click a "Check Answer" button to improve his or her learning results on the fly until the right answer is matched in the database. Moreover, improvements in Learning Journey make it possible to provide the learners with a detailed statistical analysis of their learning progress.

Learning Journey in turn is a performance optimizing tool for providing learners and facilitators with feedback. Winther-Nielsen and other facilitators are using it in combination with Bible Online Learner to keep track of the learning progress of their students via unobtrusive surveillance (cf. Gottschalk and Winther-Nielsen 2013). With the help of various data mining algorithms and a formal IRT-based framework inspired by Metsämuuronen (2013), facilitators are able to gather information on the learning progress of a particular student. The learning data of students who are logged on in Bible Online Learner are statistically evaluated and this yields an overview of the learning progress of the students (cf. Gottschalk and Winther-Nielsen 2013). Figure 3 shows a table with the learning progress of a student.

Logbook of SALT 03_Gasy

Filename	Start at	Duration (min:sec)	Seconds per right	Correct	Wrong	Correct per Minute	Accuracy	Proficiency
BH08_1_NOUNS_Gen0101-05.3et	2014-11-28 09:18	02:36	5.4	29	4	0.34	8.25	0.3
BH11_VERBConjugation_Gen0101-05.3et	2014-11-28 09:10	07:17	10.9	40	12	0.11	4.33	0.08
1. Vocabulary 210-230.3et	2014-11-24 15:44	18:22	55.1	20	6	0.04	4.33	0.03
1. Vocabulary 210-230.3et	2014-11-24 15:28	15:36	93.6	10	15	0.03	1.67	0.01
Verbs00_Lexeme_English_Form_All Regular in Genesis 1-3.3et	2014-11-24 14:49	37:53	51.7	44	61	0.01	1.72	0
Vocabulary 25-40.3et	2014-11-07 13:00	27:50	104.4	16	14	0.02	2.14	0.01
Verbs01_ Regular.3et	2014-11-07 12:08	38:00	73.5	31	74	0.01	1.42	0

Quickoverview on student SALT 03_Gasy

You won 48.14 PLOTpoints
You lost 26.57 PLOTpoins
This results in 21.57 PLOTpoints after this training sessions!
You took 56.37 seconds on the average to give a right answer!

Right now the best advice is to work on these three activities
1. Verbs01_ Regular.3et Here 70.48% of the exercises caused problems
2. Verbs00_Lexeme_English_Form_All Regular in Genesis 1-3.3et Here 58.1% of the exercises caused problems
3. Vocabulary 25-40.3et Here 46.67% of the exercises caused problems

Your accuracy is: 2.02!
Your proficiency is: 0.65!
You answer 1.29 answers correct per minute on the average!
To individuell analysis

Figure 3 On overview of the learning progress of a student from Madagascar displaying his learning logbook and a brief overview of his progress together with a small recommender function

As can be seen, the tool uses measures like *accuracy* and *proficiency*, which are rooted in psycholinguistic research. In the formal framework I developed together with Nicolai Winther-Nielsen and which is introduced in Gottschalk and Winther-Nielsen (2013) and Gottschalk (ms), the concept of automatization is also used, which is discussed in DeKeyser (2012). The parameter *proficiency* is defined as in (1):

(1) Proficiency = sum of right answers / right answers per minute

Winther-Nielsen and Hery (2013) have shown that a low response time is not a perfect indicator for successful learning. The reason for this is that some students can achieve a high speed in answering the quiz questions while at the same time submitting a very high number of wrong answers. This is the reason why Winther-Nielsen and Hery suggest calculating proficiency as in (1) (cf. Gottschalk ms). *Accuracy* is calculated as in (2) below:

(2) accuracy = (sum of right answers + sum of wrong answers) / sum of wrong answers

What this parameter shows is the degree to which learners have responded correctly compared to the degree of wrong answers (Gottschalk ms). It is the goal of Learning Journey to give facilitators and learners the possibility of monitoring their learning statistics through self-monitoring. Peers and teachers can keep track of the learning progress through surveillance. According to the theory of persuasive technology in Fogg (2003: 46), surveillance is defined as: “One party monitors the other party to modify the behavior in a specific way”. Surveillance is a common technique when persuading people to change their behavior (cf. Gottschalk and Winther-Nielsen 2013). Via surveillance, the learners receive corrective feedback relative to their practice. Feedback, as has already been pointed out in this section, is what learners using computer-assisted language learning desire, and what enables the teachers and learners to improve the learning experience (cf. Gottschalk and Winther-Nielsen 2013).

The surveillance system is organized like a logbook of a journey that documents the various steps in reaching a goal. It contains a brief overview, with an overview of the learning progress of a student as shown in figure 3 above.

In the brief overview, how many PLOTPoints a student has won and lost is displayed, and it is

set out how may PLOTPoints this results in. Originally the idea why PLOTPoints are added to Learning Journey had to do with the idea of adding gamification to the tool. This quick journey provides insights into how long it took a student on average to give a right answer, and it has a small recommender function which gives advice on what the student needs to prioritize for practice (cf. Gottschalk ms). Furthermore, the quick overview contains accuracy, proficiency, and the number of right answers a student gives on average. Also, the quick overview contains information on the IRT model developed for a student containing data on his *ability* to master Biblical Hebrew as well as the likelihood with which he will respond correctly to the next set of exercises in Bible Online learner (cf. Gottschalk ms),

I will not discuss all details of Learning Journey. Instead, they can be found in Gottschalk and Winther-Nielsen (2013). Nevertheless, I wanted to give a flavor of what the system does. Winther-Nielsen can use it in his facilitation, and, to support his students in their learning, he analyses the statistical data in Learning Journey Online on a regular basis and gives learners feedback about what they can improve on. Beside the report of learning data presented in figure 4, Learning Journey also contains numerous more detailed analysis of learning progress, providing Winther-Nielsen with a substantial and wide-reaching overview of a student's learning progress.

In principle, students can also investigate their learning progress themselves but since Learning Journey currently does not have an extensive role and a permission model, this opportunity is not used.

Meanwhile, earlier analyses done within the EuroPLOT project have shown that an effective persuasive learning method is to give the learners the possibility to learn through active exploration. Winther-Nielsen describes this approach as follows:

> Learners must begin with an overall understanding of the form of the word before they can be persuaded to learn actual forms. This enables the learner to gradually acquire a higher level of mastery because the learner is in complete control of planning the content, pace, and range of his or her learning journey through the interface. (Winther-Nielsen 2013: 5)

An example for how the learning content in Bible Online Learner can be actively explored is displayed in figure 4 below:

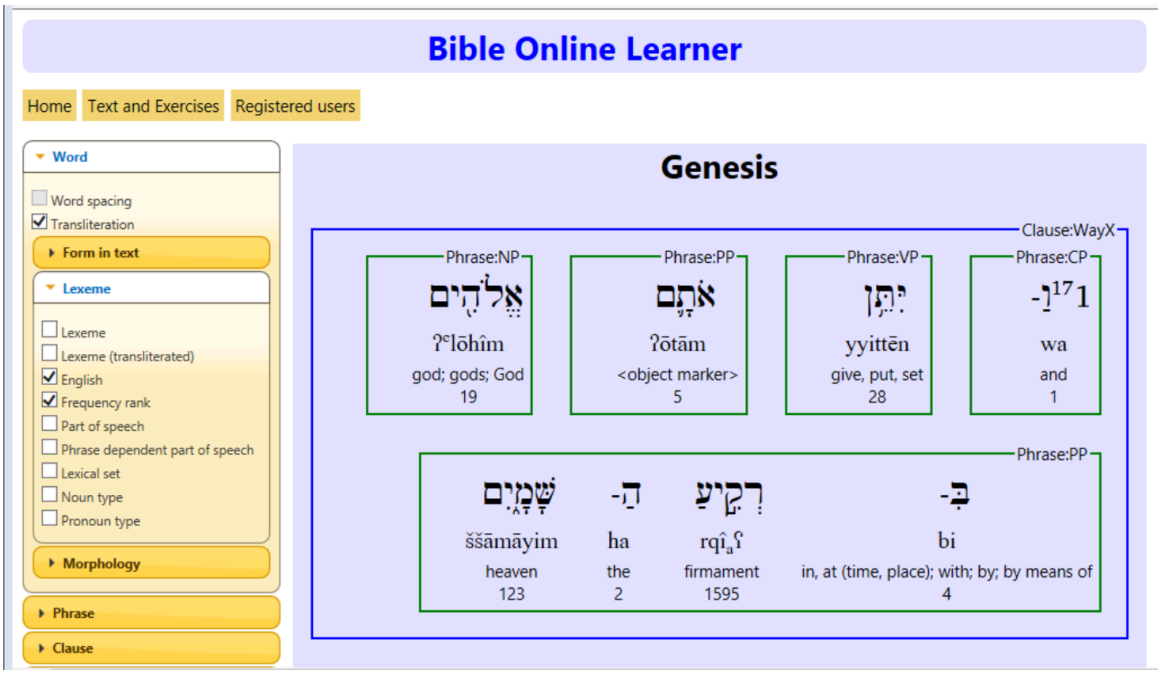

Figure 4 The active exploration of the ETCBC corpus in Bible Online Learner on Genesis.

When the learner interactively explores the corpus, the conceptual learning process (CL) in Laurillard's model in figure 2 is engaged (cf. Winther-Nielsen 2013a). The model action (BE) has been prepared through the conceptual input of the facilator. This is done for the explanation of concepts (CT). With this feature of Bible Online Learner, the relevant content of the ETCBC database is accessible through pop-up displays in a user-friendly, mouse-over fashion (cf. Winther-Nielsen 2013a: 7).

Finally, Bible Online Learner enables contextualized language learning as it is connected with a picture database displaying pictures emphasizing the current content of the biblical text. As Winther-Nielsen (2013a: 7) explains: "The ultimate goal is of course to let the learning technology generate scaffolding that activates content for the Vygotskyan zone of proximal development."

Nicolai Winther-Nielsen and Susanne Lubago, who is his student, experienced how tricky computer-supported vocabulary learning nevertheless is, even if one uses Bible Online Learner, or, as Winther-Nielsen does, provides the students with the possibility of using tools like Bible Online Learner or Learning Journey. Winther-Nielsen's experience with supporting

his students in vocabulary learning with Bible Online Learner is although with Learning Journey he can very well supervise the students as he receives a detailed statistical analysis of their learning progress, can exactly pinpoint their problems and is able to provide them with detailed feedback the students in Copenhagen did not accept the tool. Instead, the students used Memrise, a learning platform where they can generate quizzes on their own and use an endearing gardening metaphor as a gameful way to motivate themselves to use the platform.

Clearly, Bible Online Learner in its current design does not look as mature and as endearing as Memrise, and its usability is not as smooth in some places, but Winther-Nielsen did note something interesting when he did a vocabulary test - the few students who *had* nevertheless used Bible Online Learner to practice Biblical Hebrew had more success in the test, than students who used Memrise, who had problems in applying the vocabulary, which they had been learning in decontextualized quizzes in Memrise. Interestingly, a control group in Madagascar, which is supervised by Christian Højgaard, is making significant progress in learning Biblical Hebrew just with Bible Online Learner.

Sunsanne Lubago is participating in Winhter-Nielsen's Hebrew class as a distant student and was one of the few students who has been using Bible Online Learner right from the start. Her experience with the tool is that it is supporting her in exploring the Hebrew Bible. The mouse-over function in Bible Online Learner, which displays the glossing of a word helps her to learn Hebrew by reading the Bible and also the grammar drills in Bible Online Learner are fun but the problem is that the same exercises are repeated too often.

Susanne explained that she can best learn vocabulary by reading a text and Bible Online Learner is providing her with a fully glossed corpus. Nevertheless, learning Biblical Hebrew in Winther-Nielsen's class is a fulltime job in its own and beside her day work as a teacher it was very difficult for her to not keep pace with the other learners and the speed in which Winther-Nielsen discusses the different grammatical phenomena in class.

Susanne is watching the video lectures Winther-Nielsen updates on the Moodle learning environment, is in regular contact with Winther-Nielsen who encourages her with her good results which he receives from Learning Journey and she even has a private teacher. Nevertheless she did not pass the first test. "The problem is that Bible Online Learner is a tool which supports reading the Hebrew Bible but the vocabulary test, we had to pass asked for

words out of context. This made it difficult for me", explains Susanne. The test is generated by the Norwegian head quarters of the FIUC and Winther-Nielsen does not have any influence on it. It will be a future task for Winther-Nielsen to convince the administration in Norway that exams must be adjusted to teaching methods. The promising data Christian Højgaard, his teaching assistant, collected in at SALT in Madagascar will be very supportive in this dicussion.

At the moment, there is no valid statistical data for why the other students in Copenhagen do not use Bible Online Learner to the extent that Nicolai Winther-Nielsen wishes, but one hypothesis is that the user experience, usability and screen design of Bible Online Learner makes students reluctant to use the tool. The students in Madagascar are new to computer-supported language learning via a network and hence they cannot really compare Memrise and Bible Online Learner and so they accept the tool the way it is, with its primitive design. The result is that, compared to the students in Copenhagen, the group in Madagascar is making better progress towards their learning goals.

2 How to teach Biblical Hebrew: Approaches now and then

If one seeks to understand how the tools Winther-Nielsen uses in his Biblical Hebrew class work from a theoretical perspective, it is important to understand how Biblical Hebrew has been taught now and in the past within the long tradition of teaching this ancient language. Following Quast (2009) citing Ellis (1998), there exist four options concerning how grammar can be taught and within the various methods of teaching students to acquire Biblical Hebrew, these options are used in different ways:

(3)

a) input
b) explicit instruction
c) production practice
d) negative feedback

Ellis (1998) has illustrated this in the figure 5 given below, which shows how the various options for teaching Hebrew interact, and, consequently, follow each other:

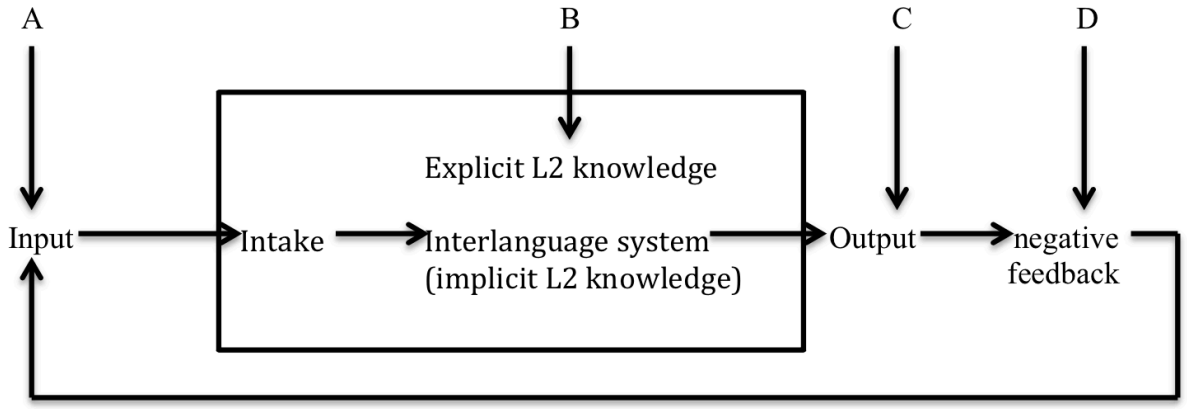

Figure 5 The various options to intervene in grammar teaching: Input (A), Explicit L2 knowledge (B), Output (C) and negative feedback (D)

One way to intervene in the process of language learning is for the facilitator to give learners the possibility of being exposed to authentic examples of the target language (cf. Quast 2009). As has been shown in section 1, Winther-Nielsen does this when he encourages the learners to explore the ETCBC corpus displayed in Bible Online Learner as, this way, the attention of the students is drawn to the target structure of Biblical Hebrew in the Bible. A method like this is called input enhancement, and indeed the input is enhanced with Bible Online Learner as the Bible is used as an interactive tutor for the learners. Another option for learning via enhanced input is to structure input activities by manipulating the language input in some way. This way, comprehension of a grammatical structure in the language in question is facilitated (cf. Quast 2009).

What teaching Biblical Hebrew for more than thirty decades has shown Winther-Nielsen is that manipulating the input for the learners is a helpful way to prepare them to derive lexical rules, which enables them to properly translate from the Bible. This is the reason why he developed a pedagogical way of displaying the semantics of verbs like נתן *nātan* properly within Bible Online Learner, and why he developed a learner-friendly approach within the theory of Role and Reference Grammar (cf. Winther-Nielsen ms).

A second method of teaching learners Biblical Hebrew is explicit instruction (cf. Quast 2009). In a typical Hebrew lesson given by Winther-Nielsen, he provides the students with explicit information on the target structure of the language and, additionally, he provides the students with short videos about specific phenomena in Biblical Hebrew which they can download from the Moodle platform he provides for his students (cf. Quast 2009). It is also

possible to make the learners analyze a collection of data illustrating a particular grammatical phenomenon. This way, the student is enabled to derive a set of rules from the data, which explains the phenomenon (cf. Quast 2009). This is a technique which is not used in the current version of Bible Online Learner. The students do grammatical drills in order to learn the language by solving quizzes and they explore the corpus of the Hebrew Bible, but they do not investigate specific grammatical phenomena by using the corpus.

This will change when the vocabulary learning strategy presented in this master's thesis is implemented. Here the concept is the following: The learners have to answer questions regarding the semantics of a Hebrew word in order to be able to type in the translation of that Hebrew lexeme. They use an interactive dialogue to do so, and by using it they use implicit instruction.

Furthermore, a teacher can make learners produce the target language (cf. Quast 2009). This is a method not used in Bible Online Learner. The reason for this is that Nicolai Winther-Nielsen's opinion on translating from e.g. English or Danish into Biblical Hebrew is that it makes little sense. From his perspective, knowing Biblical Hebrew means being able to translate from Hebrew into English or Danish but not vice-versa, as this is a rather unnatural way to use an ancient language. Therefore, this strategy is not propagated in Bible Online Learner. However, a final method, negative feedback, is used extensively in Bible Online Learner: Users of the tool complete grammar quizzes, and they immediately receive feedback from the system if they complete the quiz correctly or did wrong.

With Learning Journey, this method is developed even further: Christian Højgaard, who teaches Biblical Hebrew at SALT in Madagascar, reported the following story in an email:

> [...] A few days ago I showed our best Hebrew student how to watch his own statistics in your Learning Journey. He saw the high score and realised that other students at SALT had more proficiency points than him. Since then he had practically speaking been working ever since which means for hours every day. Yesterday he worked from 2-11pm and even today he is working. I guess he will top the high score list in a few days :) His name is Valisoa so you can see for yourself [...] (excerpt from an e-mail from Christian Højgaard from 2014-11-30).

What Christian Højgaard's story about language learning with Bible Online Learner and Learning Journey shows is that the tools are a persuasive means of supporting learners in improving their language skills. They motivate learners to practice and become better learners.

I have shown so far how the different methods in teaching Biblical Hebrew are employed in Bible Online Learner and Learning Journey. Now the question is how these methods are used in the didactics of Biblical Hebrew, and how they are employed in Bible Online Learner. In the overview of attested teaching methods, I will focus on modern methods, which are not older than approximately 150 years old. Although Biblical Hebrew as an ancient language has been taught for many centuries, old Rabbinic methods from the middle ages do not apply these days, and in the context of computer-supported language learning, they are of no use.

A very popular method in the didactics of teaching Biblical Hebrew is the Grammar-Translation Method (Quast 2009). The focus of this method is on teaching the grammar of the language explicitly to the student and then testing their knowledge of the language in translation exercises. This method was typically used to give students access to literature in languages like Latin and Greek even though these languages are now only used scholastically and ecclesiastically. This method is still very popular when it comes to teaching Biblical Hebrew. (cf. Quast 2009).

When Winther-Nielsen teaches his students Biblical Hebrew in his class in Copenhagen, he also uses the Grammar-Translation Method to some extent. Bible Online Learner supports the students in learning this foundational, exegetical method, and in the final exam at the end of the semester, students are asked to translate a passage from the Hebrew Bible into Danish.

What supports Winther-Nielsen in using the Grammar-Translation Method in his classes is the use of short videos, which the students can download from Moodle. This way, he uses explicit instruction, the second teaching method mentioned above. This strategy enables the students in learning a set of grammar rules as well as in using a specific set of grammatical terms (cf. Quast 2009). Winther-Nielsen presents the language of Biblical Hebrew to his students in a deductive and systematic way, and each lecture focuses on a specific grammatical phenomenon e.g. irregular verbs (cf. Quast 2009). Bible Online Learner enables the students to systematically practice the grammatical phenomena that they encounter in their Hebrew lessons.

It is the goal of the Grammar-Translation Method to convert the knowledge the students have in the L2 they are acquiring into an L1 (cf. Quast 2009). When this method is used in a

traditional context, this is done by making a dictionary and other reference tools available to the students to make them read and translate the Hebrew Bible. By using Bible Online Learner, something has changed: The students can use a fully glossed and transliterated interactive corpus of the Bible to train their reading abilities; both Winther-Nielsen (ms) and Gottschalk (accepted) envision that the fully glossed ETCBC corpus will be adjusted in such a way that only an exactly matching gloss for words like נתן *nātan* is used in the ETCBC corpus.

Teachers like Winther-Nielsen, who use the Grammar-Translation Method most of the time, talk with their students *about* Biblical Hebrew, they do not talk *in* Biblical Hebrew (cf. Quast 2009), but, as noted before, there is good reason to not teach the students to talk in Biblical Hebrew. Since Biblical Hebrew is an ancient language, there is only little known about its phonetics and phonology and, additionally, it is rather difficult and pointless to speak in an ancient language without a vocabulary which is adjusted to modern needs.

The approach of the inductive method for teaching Biblical Hebrew applies for advanced students of the language (cf. Quast 2009) and, although it could perfectly well be supported with Bible Online Learner, neither Winther-Nielsen nor any of the other Hebrew teachers I have interviewed while working on this master's thesis have applied this method.

Behind the name *Inductive Method* is the idea that the teacher discusses grammatical phenomena with the students when they encounter them in a text, e.g. in the ETCBC corpus stored within Bible Online Learner. This means the Inductive Method has an emphasis on reading larger amounts of authentic text (cf. Quast 2009).

While the inductive method can be realized with ease using a tool like Bible Online Learner, it is a fact that methods which make use of incidental reading can only be successful when accompanied by methods using explicit vocabulary learning, as knowing a relatively huge amount of words is necessary for being able to understand a text written in Hebrew without great difficulty. I cannot give an exact number here as the literature differs considerably when discussing the point of the exact amount of vocabulary. Of course, the fact that the corpus used by Bible Online Learner is glossed makes incidental reading easier, as the student can use a simple 'mouse over' to receive a translation of a word, but, as pointed out in Winther-Nielsen (ms) and Gottschalk (accepted), there are several translations for one word in the

corpus and merely knowing the words does not help the student if he does not have any knowledge of Hebrew morphology and syntax. This is the reason why Winther-Nielsen refrains from using the inductive method in his classes.

The following three teaching methods are not used in any of the Hebrew classes I have investigated for a reason mentioned quite often in this section: they require the actual *use* of Biblical Hebrew in the sense of production and / or comprehension. The Audio Lingual Method focuses almost exclusively on production practice and negative feedback. Here, the student talks with his teacher in the target language and receives negative feedback in the case of grammatical errors (cf. Quast 2009). In Communicative Language Teaching, there is no focus on the mastery of structures, but rather on the communicative potential of the language. The student communicates in the language within real-world tasks which the teacher develops for him (cf. Quast 2009). Finally, the idea of the Total Physical Response method is that the students acquire the language very much in the same way as they learned their native language. Here, they use their bodies much the same way as a child responds to instructions from their caregiver, and, essentially, the students follow instructions from the teacher and move their body accordingly (cf. Quast 2009). None of these methods are realized in Bible Online Learner. The reasons for this have already been raised in the preceding text.

The final method to be discussed is Task-Based Language Learning: The approach in task-based language learning is that learners have to use the target language they are acquiring to achieve meaningful goals. The foreign language is learned in real tasks ranging from getting directions in a new city to ordering food in a restaurant (cf. Hooper et al. 2012). Task-based language learning in this way is of course not possible with Bible Online Learner but when learners like Susanne Lubago solve the quizzes in Bible Online Learner Winther-Nielsen has generated for her it is similar to solving little tasks to acquire a language and from this perspective learning Biblical Hebrew with Bible Online Learner is also motivated by task-based learning.

3 Vocabulary and strategies for acquiring it

Now that I have written in depth about how Biblical Hebrew is taught both generally and specifically in Nicolai Winther-Nielsen's class, the question is what vocabulary is and how it can be best acquired, i.e. which strategies exist for supporting learners in learning new words

in a foreign language? What does the literature say about how Susanne Lubago could have facilitated the acquisition of vocabulary for her Hebrew class in the best way?

Generally, an accepted definition of vocabulary is to say that vocabulary consists of lexical items or lexemes. A lexeme can be said to be a unit of meaning belonging to a set of grammatical forms. That means the verb *run* with its forms *runs* and *running* can be said to be a lexeme with various inflectional forms. This definition is in accordance with Jackendoff (2003) and Pinker (1994), who note that words formed by a regular pattern of inflection should not be considered as separate items in the mental lexicon, while irregular forms and semi-productive forms like *run – ran* should be regarded as lexically different forms.

Choosing a word like *run* already shows the dilemma of defining vocabulary: While *run, runs, running* are regular forms and can be regarded as belonging to the lexeme *run,* what should be said about the irregular form *ran*? Is *ran* a vocabulary item of its own? What do we do about the derived form *runner*, or with the conversion of *run* from verb into noun? What should we say about proverbs such as *Still waters run deep*? In this proverb, the whole is more than the sum of its parts, as water does not normally 'run' like a human.

Thompson (ms) deals with this dilemma as follows: vocabulary may be defined as independent units consisting of individual words (*lemmas)*, but also as multi-word units, irregular forms, semi-productive forms, and derived forms (cf. Thompson ms).

When I talk about strategies for how a learner can best acquire vocabulary, I mean that by using Bible Online Learner, the learners in Winther-Nielsen's class acquire any aspect of knowledge about a lexical item (e.g. its meaning(s), its written from, its spoken form, its grammatical behavior, and its collocations, associations and frequency) in such a way that it can be passively recalled by carrying out exercises with Bible Online Learner (cf. Thompson ms).

In my use of the term *vocabulary-learning strategies*, I follow a particular approach by Thompson (ms). He uses vocabulary-learning strategies as a means to enable vocabulary learning in a learner. For him, two leading questions are what vocabulary learning strategies are and which vocabulary learning strategies would be most helpful for learning an amount of Biblical Hebrew vocabulary which is suitable for reading the Old Testament. He asks the

question of what vocabulary learning strategies are because they are not clearly defined in the literature. Thompson (ms) explains that there is no consensus in Applied Linguistics about an adequate definition of them. Nation (2001: 217), for example, states that: “It is not easy to arrive at a definition of what a strategy is ...”. Moreover, Schmitt (1997), who carried out some research on vocabulary acquisition and vocabulary learning strategies, proposes no explicit definition, although he discusses possibilities for vocabulary learning strategies in great detail (cf. Thompson ms). Thompson himself lists the following categories for strategies of vocabulary learning:

(4)

- strategies involving authentic language use
- strategies involving creative activities
- strategies used for self-motivation
- strategies used to create mental linkages
- memory strategies
- visual / auditory strategies

  (cf. Thompson ms)

For Thompson, all these strategies are various means for enabling a learner to acquire vocabulary. A strategy in this context is something like a *technique* (either mental, physical, or audio / visual) that makes learning vocabulary easier for a learner. Like Thompson, I use vocabulary-learning strategies in my design approach for improving vocabulary learning with Bible Online Learner and Learning Journey. Of special interest in this design approach are learning strategies which support or are supported by corpus-driven language, and incidental and intentional reading.

Due to the specific characteristics of Biblical Hebrew vocabulary, not all vocabulary-learning strategies listed in (4) are successful. As already noted in section 1, Biblical Hebrew is a text-based language. In so far as the phonetics of Biblical Hebrew are known, they limit the efficacy of auditory learning strategies such as listening to recordings of the language because of similarities in the pronunciation of several letters in the Hebrew alphabet (cf. Thompson ms). Like Winther-Nielsen, Thompson (ms) abandons all learning strategies for Biblical Hebrew which have to do with attempts to *speak* the language. When it comes to the use of cognates in learning Biblical Hebrew, vocabulary learners must recognize very early on that

there is a genealogical distance between Biblical Hebrew and their own languages. Biblical Hebrew is a Semitic language, and English is a Germanic language, for example. Cognates like *brother* in English and *Bruder* in German do not exist between Biblical Hebrew and English and, consequently, vocabulary-learning strategies involving cognates are condemned to failure (cf. Thompson ms; Schmitt 1997).

Several studies (Gu and Johnson 1996; Ponniah 2011; Webb 2008) have shown that actively reading in an L2 language enables learners to develop large vocabulary sizes. This is the reason why vocabulary-learning strategies using authentic language are extremely important learning strategies (cf. Thompson ms). While Winther-Nielsen of course cannot confront his learners with authentic language use in the sense of making his students read a newspaper, he can make the students in his class read Biblical Hebrew. As already mentioned, the possibility of actively exploring the ETCBC corpus with Bible Online Learner is such a technique and, in the second paper which is part of this thesis, I will specifically discuss it to show how it can be used to improve vocabulary learning with Bible Online Learner and Learning Journey. Generally speaking, to already give an impression at this stage of the study, findings in the literature are that extensive reading of literature as a strategy for vocabulary learning can be a form of support for language learners, if they are also supported by explicit learning (cf. Ellis 1994; cf. Thompson ms).

One of the learning strategies mentioned by Thompson can roughly be described as 'us[ing] a computer program to practice words', and, indeed, this strategy seems to be effective in the case of learning second language vocabulary, as tools like Anki (http://ankisrs.net) or Memrise (http://www.memrise.com) promise. Anki is a tool that provides its users with digitally editable flash cards. Users can define decks with several hundred digital learning cards, which they can then use to memorize facts and vocabulary. In the course of a learning session, the user clicks through a deck and decides how long it should take until a specific card is shown again. If a user knows a word very well, the time between him initially being shown the card next repetition can be extended. In Memrise, on the other hand, a user can construct quizzes for himself. Thus it is, for example, possible for a user to ask 'What does the verb נתן nātan mean?', and then up to four possible translations can be given. The user can click on the solution which he thinks is right, and is immediately shown the correct answer. Memrise uses the metaphor of gardening and uses gamification to motivate the user to stay with the tool and keep learning.

In the area of applied computational linguistics, there exist several studies examining the use of computer programs for learning vocabulary. Thompson gives quite an extensive overview of these studies and I paraphrase his literature review here. Some studies make students read text from a corpus. Here, Chun and Plass (1996) found out that students who read a text on a computer screen are able to learn the vocabulary *if* they have the possibility of looking up the words easily, or if the words in the text are glossed with both text and pictures (cf. Thompson 2011). The success of this method is that it leads to both good short-term and long-term retention. Following Lyman-Hager (2000), glosses enhance the comprehension of a reader, especially if the text contains a high incidence of unknown words. Leffa (1992), on the other hand, has investigated the use of hypertext technology using an electronic glossary for developing reading comprehension. This study has shown that an electronic glossary is more efficient than a traditional dictionary. An electronic dictionary allows the subject to understand 38% more of a passage in less than 50% of the time (cf. Leffa 1992; Thompson 2011). Thompson (ms) sums these findings up and explains that students using computers for reading begin reading much earlier, and begin to gain aspects of vocabulary knowledge which are better acquired through implicit learning.

When it comes to the use of computers and vocabulary learning, many researchers have shown that the use of computers has the advantage of contextualization (cf. Thompson ms). It has been pointed out several times that learning words using methods such as word lists and flash cards (as in Anki and Memrise) which exist outside an authentic context causes problems, and this when it comes to the active as well as passive use of acquired vocabulary (cf. Nagy 1997; Thompson 2011). Thompson cites three studies showing that computer-supported language learning offers good opportunities for a hybrid approach between learning vocabulary lists and contextualized learning. While Cobb (1999) and Groot (2000) focus more on advanced learners, Kang (1995) suggests that computer-supported vocabulary learning using contextualization is a promising approach for beginning language learners. In his approach, Kang tested the effects of using a computer to introduce words within their genuine context and was able to show that presenting words along with both written and visual contexts was much more successful for definition recall, listening comprehension and knowledge transfer than paired-associate learning using just paper and pencil (cf. Kang 1995; Thompson 2011).

Some of Winther-Nielsen's students have used Memrise to facilitate vocabulary learning, and when the students using this method had to solve vocabulary problems, the students failed even though they had good results in Memrise. Although we have no scientifically valid data concerning this incident, it seems to be the case that learners who learned vocabulary in context (like Christian Højgaard's group at SALT in Madagascar which uses Bible Online Learner very actively) are more successful in vocabulary tests. I will come back to this issue in detail in the second paper, which is part of this thesis.

4 Problem statement and outline of this thesis

In a personal analysis and evaluation (attached to this thesis in the appendix) of teaching Biblical Hebrew at FIUC in Copenhagen Nicolai Winther-Nielsen reflects on the problem that his students have not accepted Bible Online Learner as a tool for learning. He comes to the conclusion that in the first two weeks of the semester he has spent too much time on lecturing on verb stems instead of focusing on an important midterm test on the most frequent Biblical Hebrew vocabulary. At the same time the students started using Memrise for vocabulary learning to pass the test and did not use the videos on grammatical phenomena Winther-Nielsen has provided them with and skipped class.

The class was scared of the midterm test in advance and 50% of the class failed the test which resulted in more students focusing on using Memrise for a preparation of the final exam and less students using Bible Online Learner. Also many students stopped going to class and they did not prepare for the class when they attended it, which made teaching even more difficult. In a second test in October again two good students failed in the vocabulary test. One failed because it was hard for him to learn with the glosses in Bible Online Learner and one because he did not have sufficient time for practicing Hebrew. A third student simply has general problems to learn Hebrew and failed in the exam. A third test is passed by an online student and another student fails again. Finally in December only two gifted students of Winther-Nielsen are able to follow the crucial training on Hebrew syntax in class. The final exam took place on December 18th and still needs to be corrected.

In his evaluation of the course Winther-Nielsen comes to the conclusion that most of the students have not made enough time for practicing Hebrew. His course is developed for a daily workload of 5 - 6 a day and most of the students in his class have not been able to meet

that goal. On the other hand the students prefer enforcement and punishment in their classes rather than persuasion; the students in Copenhagen have not been motivated enough to practice for so many hours as self-directed learners and rather returned their responsibility to learn as self-directed learners back to the teacher from whom they expect to make them feel bad if they to do not perform well. Possibly the punishment the students asked for could have been resolved if Bible Online Learner had used gamificaion to make the learning more fun.

Nevertheless Winther-Nielsen thinks the most crucial point, which caused the high fail-rate in the course was the decontextualized vocabulary learning with Memrise. Using decontextualized vocabulary learning caused the students to not be able to derive lexical rules to determine the meaning of words like נתן nātan in their morphosyntactic context and hence prevent them to learn the vocabulary properly.

Another group of students in Madagascar taught be Christian Højgaard had more success. In his class 39 out of 49 students passed the exam. Højgaard's class consisted of two groups of students: A minor group consisted of students knowing English and being familiar with computers which meant that they could easily use Bible Online Learner and also used it a lot. The major group had troubles using Bible Online Learner because they were not familiar with neither computers nor the tool. In general, the students using Bible Online Learner had far better results than the students not using the tool. Generally however the students did not use decontextualized vocabulary learning with Memrise which can give an inside in why the students in Madagascar performed better then the students in Copenhagen.

Based on the experiences Nicolai Winther-Nielsen made during his teaching, this master's thesis deals with is the following problem: When learning Biblical Hebrew, what causes most learners serious problems is remembering and correctly applying the huge amounts of vocabulary. Many lexemes have several different meanings and, when translating from the Hebrew Bible, it is nearly impossible for the students to choose the meaning which correctly fits in the context. How can this problem be solved with Bible Online Learner and Learning Journey?

The goal of this thesis is to cure Bible Online Learner's Achilles heel which makes vocabulary learning with the tool so difficult: Namely that while the tools gives the learners to

possibility to explore the corpus of the Hebrew Bible to learn vocabulary in context, the exercises which the learners use to practice vocabulary is a decontextualized list of words with many repetitions which prevents the learners from acquiring lexical rules to derive a translation of words like נתן nātan with several meanings. As explained in the summary, this master's thesis consists of two publishable papers: My paper for the proceedings of the Role and Reference Grammar Conference 2013 is the theoretical foundation to my approach to computer-supported vocabulary learning. It deals with the question of how computer-supported vocabulary learning can be improved by using a computationally adequate model of RRG. It suggests conceptual graphs (cf. Sowa 2000) as a new approach to semantics in RRG to develop a computationally tractable version of RRG which can be implemented in a semantic parser as an extension of existing learning software for Biblical Hebrew. The application of conceptual graphs has the advantage that computational approaches for ancient languages such as Biblical Hebrew can be developed. In this approach, a linking algorithm from syntax to semantics is reduced to a set of lexical rules which match attribute value matrices defining the layered structure of the clause against an ontology (cf. Gottschalk 2010; Gottschalk 2012a; Gottschalk in press), and uses an algorithm for the automatic determination of Aktionsarten for Biblical Hebrew developed by Winther-Nielsen (ms). With this semantic parser it is possible to derive glossings for Bible Online Learner which use exact meaning of a Hebrew word depending on their morphosyntactic context. The current problem with the glossing in the corpus Bible Online Learner is using, is namely that several Hebrew words have more than one meaning and that their means are all displayed in the glossing in Bible Online Learner.

This way it is difficult for the student to learn Hebrew vocabulary by reading the Bible with Bible Online Learner as they learners cannot easily derive lexical rules which help them to determine a proper translation of a lexeme which has multiple meaning depending on its context. My hypothesis 1) regarding computer-supported vocabulary learning is that words are best learned within the morphosyntactic context they occur and to motivate the learners to explore the hypertexed corpus of the Hebrew Bible by themselves. The idea regarding an improved glossing in Bible Online Learner is: If the learners are only exposed to ideal glossings and to perfect examples they will be able to derive lexical rules for the determination of the meaning of words which have multiple meaning depending on the morphosyntactic context they occur in.

It is the task of the paper I have written for the proceedings of the RRG conference 2013 to develop a formal model within RRG to enable learners to deal with verbs like nātan, which have several meanings depending on the morphosyntactic context they occur in and to develop a theoretical approach to a semantic parser within the theory of RRG, which enables a cleaning of the corpus Bible Online Learner uses.

In the second publishable paper, which is part of this thesis, I am dealing with the question of how the learning with Bible Online Learner and Learning Journey can be improved towards computer supported vocabulary learning and how the theoretical findings from the first paper, Gottschalk (accepted), which is part of this master's thesis, and Winther-Nielsen (ms) can be practically applied within the tools to improve facilitation and learning? The second paper presents a design on a practical approach to how learners using Bible Online Learner will be able to acquire lexical rules to determine the exact meaning of *nātan* in the corresponding morphosyntactic context. It will one the hand show how exercises in Bible Online Learner needs to be designed in order to enable corpus-driven language learning and on the other hand I will show how the sound learning statistics in Learning Journey can be used for vocabulary learning. Here my hypothesis is the following: While tools like Anki and Memrise use specific-fixed algorithms which determine when recently acquired vocabulary needs to be repeated in order to be learning (cf. Edge et al. 2012), Learning Journey with its statistical framework can offer a much more flexible approach which is user-driven and enables tailoring and tunneling, which are two persuasive means, which following Gottschalk and Winther-Nielsen (2013) should enable successful language learning with Bible Online Learner. In this paper I am going to discuss the algorithms used in Anki and other tools and compare them with the approach chosen in Learning Journey. More important however is my idea that learning the vocabulary of Biblical Hebrew is best possible within contextualized vocabulary learning as done with Bible Online Learner. I present a study made with Nicolai Winther-Nielsen's class in Copenhagen and am able to show that those students who learned the vocabulary of Biblical Hebrew contextualized with Bible Online Learner were more successful in learning the language than those students who learned vocabulary decontextualized with Memrise.

Instead of providing the reader now with an overview of the structure of this thesis I will instead warn her that the unusual organization of the thesis with two publishable papers rather than one monolithic work comes with a cost: 1) I will not give the structure of the thesis here

by referring to the various sections of it as what follows are two stand-alone papers which already describe their structure in their first section by themself. 2) For each paper the numbering of sections, examples, figures and tables start with 1 which results in at least three figures 1 in this thesis. 3) There are overlappings between the introduction, the first and the second paper. The reason is, since the three parts are designed as being able to stand on their own, I cannot use cross-references to connect the three parts of the thesis with each other, therefore the reader will from time to time find passages especially in the two publishable papers, which she already knows from another part of the thesis. Nevertheless the whole thesis has a general conclusion at its end, as one would expect it from a traditional master's thesis. To not disturb the reading flow of the reader the references of the introduction, the first and the second paper are cumulated in one general section on the references towards the end of this thesis.

On the application of conceptual graphs in RRG – First steps towards a functional computational processing model

Judith Gottschalk
Aalborg University, Rendsburggade 14, 9000 Aalborg
gottschalk.judith@gmail.com

Abstract

How can computer-supported vocabulary learning be improved by using a computationally adequate model of Role and Reference Grammar [RRG] (cf. Van Valin 2005)? This paper suggests conceptual graphs (cf. Sowa 2000) as a new approach to semantics in RRG to develop a computationally tractable version of RRG which can be implemented in a semantic parser as an extension of existing learning software for Biblical Hebrew. The application of conceptual graphs has the advantage that computational approaches for ancient languages such as Biblical Hebrew can be developed. In this approach, a linking algorithm from syntax to semantics is reduced to a set of lexical rules which match attribute value matrices defining the layered structure of the clause against an ontology (cf. Gottschalk 2010; Gottschalk 2012a; Gottschalk in press), and uses an algorithm for the automatic determination of Aktionsarten for Biblical Hebrew developed by Winther-Nielsen (ms).

1 Introduction

When learning Biblical Hebrew what causes most learners substantial problems is remembering and correctly applying the large amount of vocabulary. Many lexemes have several different meanings and, when translating from the Hebrew Bible, it is nearly impossible for students to choose a meaning which correctly fits in the context. Students have various strategies when learning their vocabulary. Some write hundreds of record cards by hand with Hebrew lexemes on the front and translations in their native language on the back so that they can use them to recall the foreign words, while others use technical means. Computer programs such as Anki (http://ankisrs.net), where one can generate digital record cards called flash cards, are commonly used, or a webpage called Memrise

(http://www.memrise.com) is consulted. Here the students can learn their vocabulary in a gameful way.

In the particular approach to vocabulary learning developed by Thompson (ms), vocabulary-learning strategies are used as a means of enabling vocabulary learning in a learner. For Thompson, the leading question concerns what vocabulary learning strategies are, and which vocabulary learning strategies would be most helpful for learning an amount of Biblical Hebrew vocabulary sufficient for reading the Old Testament. He asks the question of what vocabulary learning strategies *are* because they are not clearly defined in the literature. Thompson (ms) explains that there is no consensus in the Applied Linguistics literature regarding what constitutes an adequate definition of learning strategies. Nation (2001: 217) for example states that "[i]t is not easy to arrive at a definition of what a strategy is ...". Furthermore, Schmitt (1997), who did some research on vocabulary acquisition and vocabulary learning strategies, proposes an no explicit definition although he discusses possibilities for vocabulary learning strategies in great detail (cf. Thompson ms: 24). Thompson himself lists the following broad categories for strategies of vocabulary learning:

- strategies involving authentic language use
- strategies involving creative activities
- strategies used for self-motivation
- strategies used to create mental linkages
- memory strategies
- visual / auditory strategies

  (cf. Thompson ms: 65)

For Thompson (ms), all these strategies are various means to enable a learner to acquire vocabulary. A strategy in this context is something like a *technique*, either mental, physical, or audio /visual to make learning vocabulary easier for a learner. Several studies have shown that using computers for vocabulary learning has the advantage of facilitating vocabulary learning in context. The problem with learning single words, like many students do when they use word lists or flash cards, which as pointed out above, is that it is not successful because the learners miss out on the authentic contexts in which the vocabulary occurs (cf. Thompson ms, Nagy 1997).

Thompson cites in particular the study of Kang (1995), which suggests that computer-supported vocabulary learning is a promising approach. Kang tested the effects of using a computer program to introduce words in genuine contexts and found out that presenting words along with both written and visual contexts was better for definition recall, listening comprehension, and knowledge transfer than paired-associate learning with paper and pencil, with a computer having words but no picture, or with a computer having only a picture (1995:43) (cf. Thompson ms).

Like Kang (1997), we present an approach to learning Biblical Hebrew vocabulary which uses a corpus-driven tool called Bible Online Learner (http://bibleol.3bmoodle.dk). Bible Online Learner uses the corpus of Biblical Hebrew which is provided by the Eep Talstra Center for Bible and Computer [ETCBC] in Amsterdam. Claus Tøndering has programmed Bible Online Learner in a continuation of the EuroPLOT project (http://www.eplot.eu) in close collaboration with Nicolai Winther-Nielsen who developed Bible Online Learner's persuasive design.

In the course of the extensive testing of Bible Online Learner, learners found, on the one hand, that corpus-driven language learning can be very persuasive but that, on the other hand, effective learning requires the knowledge of lexical data that is not available in the corpus (cf. Winther-Nielsen ms; cf. Winther-Nielsen 2013). The problem laid out in the beginning made using Bible Online Learner less successful for learners than one might expect in the first place. Learning the ancient Semitic language of Biblical Hebrew in a corpus-driven technology is not only a matter of being able to read and parse the Hebrew Bible. It also means being able to translate from the Bible, and translating means having an active vocabulary - not only being able to recall morphological and syntactic information in order to process the language (cf. Winther-Nielsen ms).

While Thompson (ms) suggests approaches from Applied Linguistics for improving vocabulary learning which do not use a computer, like Winther-Nielsen (ms), I come to the conclusion that a stronger linguistic framework is needed to enhance the learning technology, one which can be computationally implemented to better scaffold computer-supported vocabulary learning. Thompson (ms) thinks that computers in general do not enhance vocabulary learning given their limited availability. Nevertheless, Winther-Nielsen (2013b) has shown that computer-supported language learning can even be established in the

developing communities like, e.g. Madagascar.

Winther-Nielsen (ms) explains that using a linguistic framework as an engine for computer-supported language learning will improve corpus-driven language learning as it enables learners to derive lexical rules from the corpus. This way, learners can choose the correct interpretation of a word in their target language of translation. Winther-Nielsen shows that deriving meaning from structural valence patterns, as it is done at the ETCBC (cf. Dyk in press), is not a sufficient solution as it is not reliable in all contexts. Therefore, he suggests a solution within the framework of Role and Reference Grammar [RRG], which does not treat valence patterns as a purely syntactic phenomenon, but instead relies on semantic roles and pragmatic reference (cf. Winther-Nielsen ms).

In his most recent paper, Winther-Nielsen explores what RRG can offer in its current form and expands the theory with a solution using a WordNet ontology (http://wordnet.princeton.edu/), following a proposal made by Petersen (2007) to enrich parser output to generate conceptual graphs [CG] from it (cf. Winther-Nielsen ms). In the course of his research, Winther-Nielsen develops a pedagogical framework which enables learners of Biblical Hebrew to derive the meaning of a word in a specific syntactic context with the help of a simple algorithm that uses the theoretical basis of RRG.

In this paper, I am taking Winther-Nielsen's work up in that I develop the theoretical basis for adding a semantic parser to Bible Online Learner which will improve the automatic generation of glosses in Bible Online Learner and supports learners in their desire to derive the correct meaning of verbs from the corpus. Section 2 briefly introduces Bible Online Learner and a possible design for the integration of the semantic parser into the tool. The questions the paper deals with are as follows: 1) What does a computationally adequate model of RRG have to look like in order to enable the derivation of meaning from a corpus of Biblical Hebrew and how should RRG be revised in order to be used in a computational implementation of a semantic parser which can derive meaning from the corpus? 2) How can a lexicalist-functional approach to RRG using conceptual graphs support the development of a computationally adequate model of RRG which can be implemented in a semantic parser?

The paper is organized as follows: Section 2 introduces the background to this study and explains the motivation for using RRG to automatically derive meaning from a corpus. It

makes reference to earlier and current computational approaches to semantics in RRG. Section 3 shows how vocabulary learning with Bible Online Learner, the tool I seek to improve, works at the moment, and outlines how I seek to improve it with the model developed in this paper. In section 4, I introduce my new approach to RRG semantics using conceptual graphs based on work done by Winther-Nielsen (ms), and develop a new approach to the linking of syntax to semantics in RRG, which enables the derivation of meaning in terms of semantic structures in the ETCBC corpus.

2 Background

The motivation for specifically using RRG as the linguistic engine for the semantic parser lies in the specific architecture of RRG, which uses the term 'linking algorithm'. This use naturally results in associations of implementability for a computational device in terms of a procedurally executable processing model (cf. Gottschalk 2014). Moreover, the linking algorithm is bidirectional, and thus it links the semantic representation of a clause with its syntactic representation and vice versa. This means it is possible to derive the semantics of a syntactic structure via the linking algorithm. Van Valin (2005: 129) writes about the linking algorithm:

> Viewed in terms of a processing model, the semantics-to-syntax linking is an aspect of the production process, while the syntax-to-semantics linking is an aspect of the comprehension process. In the comprehension process, the parser would take the input and produce a structured syntactic representation of it, identifying the elements of the layered structure of the clause and the cases, adpositions and other grammatically relevant elements. It is then the task of the grammar to map this structure into a semantic representation, as the first step in interpreting it, and this is where the syntax-to-semantics linking algorithm is required. The same syntactic and semantic representations are used in both algorithms. (Van Valin 2005: 129).

It is through this strong claim by Van Valin that the linking algorithm in RRG is naturally tied to being a linguistic processing model. Van Valin uses a perspective from earlier times in computational linguistics in his architecture of RRG, and the linking algorithm perceives syntax, semantics, morphology and phonology as different and separated levels. In this system, the steps are sequentially executed as in a procedural computer program. This view led to a first attempt in the development of a formal framework for RRG which can be implemented in a semantic parsing system (cf. Gottschalk 2014).

Parts of RRG have already been implemented computationally. When it comes to syntax, Nolan (2004) presents an approach to a computational lexicon in RRG which he used

in the first implementation of an RRG-based parser. In another account, Guest (2008) developed an RRG-based parser. It uses extensions of a chart parser to parse languages with different degrees of word order. Guest executes parsing via syntactic templates instead of via rules, as suggested in Van Valin (2005). Winther-Nielsen (2009) and Wilson (2009) have developed a corpus-driven tool called Role-Lexicon Module to support linguists in their analysis of Biblical Hebrew. This system uses an EMDROS database, and implements an active chart parser to generate the layered structure of the clause. RRG has also been successfully used by Nolan and Salem (2009) and Salem (2009) to develop a system for machine translation called UniArab. Here, the linking algorithm is employed via an interlingua to bridge the translation from Arab to English. Murtagh (2011) also uses RRG for translation purposes. In this case a conversational agent using Irish Sign Language is developed which uses RRG as its linguistic engine. Nolan (2013) presents an approach which used speech acts within an agent-framework to implement RRG. Gottschalk (2012b) presents the design study of an intelligent teaching agent using RRG as a linguistic engine, while Gottschalk (2014) discusses theoretical aspects of the computational implementation of RRG from a broad perspective. In Diedrichsen (2014), an approach to an RRG-based parser for German which can handle free word order in German is presented.

Beside several approaches to the implementation of syntax in RRG, there has been extensive research on computational approaches to semantics in RRG. Winther-Nielsen (ms) gives a quite extensive overview, which I will briefly paraphrase: Mairal-Usón (2003) developed a semantic meta-language which relies on semantic primitives. The problem with this approach, however, is that it creates endless *ad hoc* lists of logical terms. The meta-language does not have a means of clearly determining what semantic primitives are (cf. Winther-Nielsen ms). Therefore, in another approach to semantics in RRG, Guest and Mairal Usón (2005) developed a solution using the mathematical notion of intervals. They tie this notion to a double ontology of predicates and objects into open-ended sets, and use distinctions from fuzzy logic. The result is yet another complex meta-language. Winther-Nielsen (2009) implemented a lexicon for verbs which employs the Functional-Lexical Module developed by Faber and Mairal (1999). Here, the problem is, however, that this lexicon has proved to be less than ideal for an ancient language like Biblical Hebrew (cf. Winther-Nielsen ms).

Periñán-Pascual and Mairal-Usón (2009) have proposed the Lexical Constructional Model. This model uses a multi-purpose lexico-conceptual knowledgebase suitable for natural

language processing called FunGramKB (cf. Winther-Nielsen ms.). In the FunGramKB aspects of RRG, semantic representations in particular are used to create a knowledge base for natural language processing (cf. Periñán-Pascual and Arcas-Túnez 2010). In the ARTEMIS project (cf. Periñán-Pascual and Arcas-Túnez 2014), the various stages in the development of a computational system employing a lexico-semantical knowledgebase are defined. These steps generate the logical structure of sentences, and demonstrate that for a computerized model of the RRG syntax-semantics interface, constructional schemas should be used (cf. Gottschalk 2014).

The FunGramKB approach, however, does not work for the Hebrew Bible (cf. Winther-Nielsen ms.) A crucial assumption within FunGramKB is that all knowledge can be formalized. The problem with respect to Biblical Hebrew is, however, that, for ancient Israel, little cultural situations are available to us. The other main problem is that access to literary portrayal of figures and events in ancient documents is possible, but only through ancient texts, so that the ontology as necessary for FunGramKB cannot be developed based on language data from Biblical Hebrew (cf. Winther-Nielsen ms.).

Winther-Nielsen's (ms) suggestion for overcoming this problem is to step back and use a much more wildly known and adaptable ontology for the lexicon. WordNet is a lexical database of English which is quite large. It has been developed at the Cognitive Science Laboratory at Princeton University (cf. Fellbaum 1998). Later, WordNet was developed into the successful www.wordreference.com from the development of EuroNet (Vossen 2002) (cf. Winther-Nielsen ms.). WordNets are relational databases for linguistic applications; they are freely available and under constant development (cf. Winther-Nielsen ms.)

In a WordNet, lexical meaning is defined from relations with other lexical units in a network of meaningfully related words and concepts; this way semantic relationships between synsets are formed and these form clusters of synonymous words (cf. Winther-Nielsen ms.). With the lexical systems in WordNets, which are created by the relation of entities within the ontology, the challenge of constructing a conceptual knowledge base as in FungramKB is overcome. Periñán-Pascual's (2013) objection to WordNet ontology is that most natural language systems fail because they do not have a sufficiently developed deep semantic knowledgebase. For this reason, he rejects the shallow semantics in WordNet (cf. Winther-Nielsen ms.)

Winther-Nielsen (ms.) argues against Periñán-Pascual and explains that because WordNet in many ways was influenced by Sowa's (1984; 2000) development of conceptual graphs, it offers a user-friendly visual representation, and it can still be implemented in computational applications. Petersen (2003, 2004, 2007) has already implemented an approach to derive conceptual graphs from the ETCBC corpus using shallow parsing and a Hebrew-English dictionary. He matches English glosses with WordNet categories and constructs conceptual graphs which represent an ontology-guided, syntax-driven and rule-based joining and refining of the graphs (cf. Winther-Nielsen ms). Wilson (2009) has implemented this work in the LEX project, where he used it for mapping Hebrew syntax into the logical structures of RRG (Winther-Nielsen 2008; Winther-Nielsen ms).

In this paper, I will, in turn, develop a computational model focusing on lexical rules which match attribute value matrices defining the layered structure of a particular clause against an ontology in order to realize the theoretical model of a semantic parser which generates conceptual graphs as its output. This, as a result, will enable the correction of glosses in the ETCBC corpus used in Bible Online Learner, and will support contextualized vocabulary learning with the tool. The work in this paper is an extension of approaches developed in Gottschalk (2012a; 2012b; 2014). It is based on Nolan (2004) and develops a new computational approach to semantics in RRG, as well as to the linking of syntax to semantics in RRG by employing Winther-Nielsen's (ms) new approach to semantics in RRG. Instead of the Dowty-based semantic representation used in RRG (cf. Van Valin and LaPolla 1997; Van Valin 2005), I employ a semantic representation using conceptual graphs as developed in Winther-Nielsen (ms; cf. Sowa 2000), and generate them via a revision of Van Valin's approach.

3 Vocabulary Learning with Bible Online Learner

Before approaches to computer-supported vocabulary learning like that with Bible Online Learner were established, learners of Biblical Hebrew used comprehensive dictionaries for grammar and translation exercises. Two things were necessary for successful translation with such a dictionary: a) a knowledge of which paradigms needed to be activated because the lexemes in the dictionary were not inflected, and b) a choice of the appropriate meaning of a verse in the Hebrew Bible (cf. Winther-Nielsen ms).

With Bible Online Learner, this task can be automated as the software can display the appropriate meaning of a word in its syntactic context. The downside of this approach, however, is that students are not forced to use their knowledge of grammar to derive the appropriate meaning of a word in the syntactic context and, hence, they rarely memorize the correct translation of a word (cf. Winther-Nielsen ms).

Extensive evaluation by Winther-Nielsen (2013a) during a course in Copenhagen has shown that better means for learning vocabulary with Bible Online learner are necessary. The current solution in Bible Online Learner is that Hebrew glosses are displayed with their frequency of occurrence in the Hebrew Bible. How this is done for the Hebrew verb נתן *nātan* is shown in figure 1.

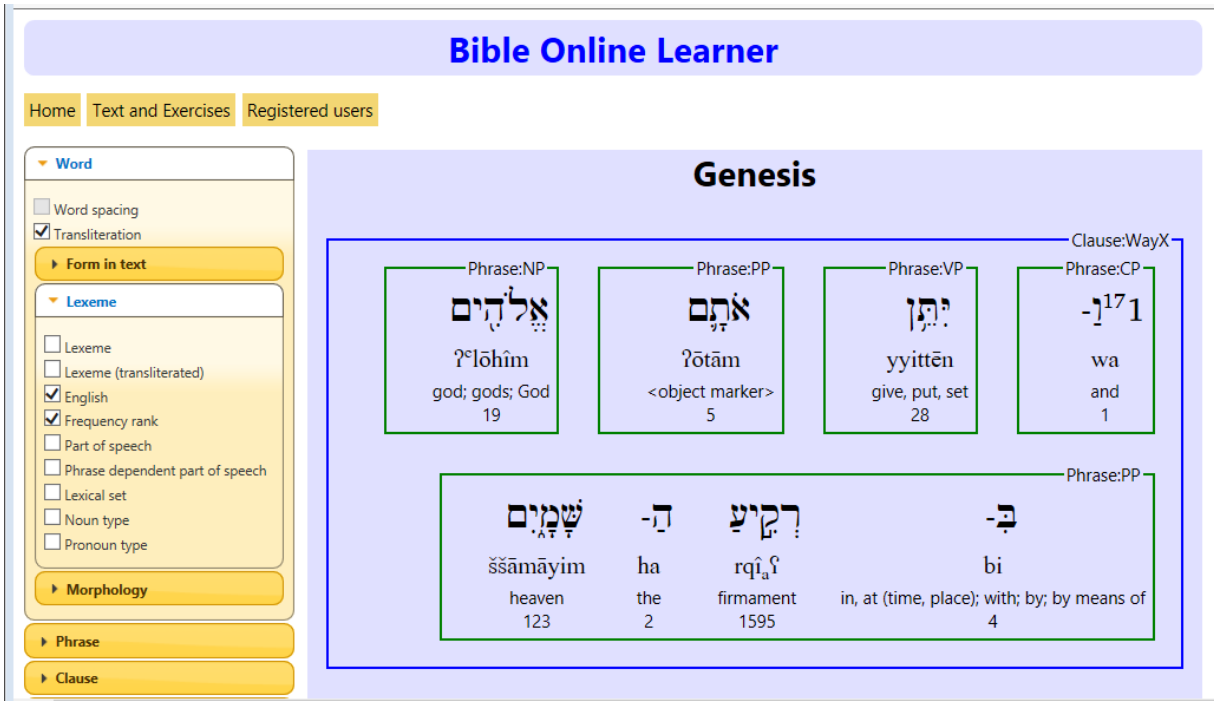

Figure 1 Display of glosses and frequency in Bible Online Learner

The idea was that the learners could explore the glosses in the corpus and practice memorization of the meaning of a word in the context of a clause. In the case of נתן *nātan*, three glosses are displayed: *give, put,* and *set*. The verb is ranked at position 28 on a list of relative frequency for all words in the Hebrew Bible. In Bible Online Learner, in turn, learners are provided with exercises based on the frequency of the words in which the various lexemes are listed and learners can enter a free translation of the word into a textbox.

The problem with the current form of exercises in Bible Online Learner is, however, that they do not support learners in choosing between one of the three meanings listed in the glosses in figure 1 (cf. Winther-Nielsen ms.). Winther-Nielsen illustrates the challenge for the learners as follows:

> To illustrate the challenge for the learner, the verb נתן *nātan* is used 5 times in Genesis 1-3. For the example from Gen 1:17 in Fig. 1, the learner has to deduce from the context that the luminaries are the referents for the masculine plural clitic pronoun attached as suffix to the object marker את *ʔet.* To 'give' them 'in the firmament' clearly does not involve an animate receiver, and therefore the two synonymous glosses 'put' and 'set' are more appropriate for placing the luminaries in visible view in the sky. In Genesis 1:29-30 נתן *nātan* has the meaning 'give' and is followed by a receiver 'to you' in a prepositional pronominal followed by the gift itself, 'all herbs and trees'. (Winther-Nielsen ms. 3)

Clearly, the learners can search for the translations in a dictionary and choose the one which might fit the rest of their translation, but this way of hunting for dictionary interpretations and their attempts to match a preferred Bible translation from their memory prevents them from learning lexical rules that they can memorize and automate for language acquisition. The current approach does not support the learners in constructing rules as part of the process of language acquisition. This is where the RRG-based semantic parser, for which I am developing the theoretical basis in this paper, comes into play. Instead of providing the learners with several translations in the glosses in Bible Online Learner, the task will be to automatically derive the translation which exactly fits the syntactic context which a verb like נתן *nātan* occurs in. If this is done, for example, in the case of the gloss in Gen 1:17, it would be revealed that *put* is a better translation than *give* using this method.

Moreover, the exercises in Bible Online Learner will be improved. Instead of providing the learners with multiple lexemes outside of their syntactic context and asking them to enter all possible translations into a textbox, the learners will be provided with text excerpts from verses in the Bible with the lexeme occurring in its syntactic context. Winther-Nielsen (ms) has developed an algorithm which makes it possible to derive the meaning of a verb from its semantic context, and the learners will be 'forced' to follow this algorithm before they can enter the meaning of a verb into a textbox. Winther-Nielsen's algorithm will be discussed in section 4 in some detail. For now, I will give simple examples of how it can be employed for learning purposes. For instance, before entering the meaning of a verb into the textbox, the learners are asked specific questions which help them to derive the meaning of the word e.g. 'Is the Aktionsart of the verb a causative?'. Here, the user can choose a radio button with 'Yes' or 'No' and is directed to the next question he has to answer: 'Does the verb have the

Aktionsart process?'. Again, the learner has to click a radio button and is, in this way, directed to the correct translation and forced to follow a clearly defined algorithm to derive the meaning of a word so that he will be able to develop knowledge of lexical rules for Biblical Hebrew vocabulary.

4 Conceptual Graphs in RRG: Towards a formal model for implementation

The way to proceed in implementing the RRG linking algorithm in order to develop a semantic parser is to design a computationally adequate formal model of RRG. When talking about computational adequacy in RRG, I mean that the theory should refer to formal systems which are computationally tractable and can be executed on a computational device in an algorithmic way in a finite time. This concept of computational adequacy supports the levels of explanatory adequacy required within a Chomskyan framework (cf. Gottschalk 2014; Gottschalk 2012a). These two levels are interconnected because a theory which is not computationally tractable and which has a Turing complete generative power is, following Carpenter (1991), who investigated HPSG, less explanatory than a theory which *is* tractable, *can* be executed on a computational device and has *less* generative power in the sense of not being Turing complete (cf. Gottschalk 2014, Gottschalk 2012a: 130).

If a linguistic theory is computationally adequate, it can process language with low storage demands without mistakes. A crucial assumption in this context is that computational adequacy is based on the Church-Turing thesis. This assumes that everything which is computable on a machine is intuitively computable, and vice-versa. The consequence of the Church-Turing thesis being bidirectional is that since a natural language is intuitively computable it should also be computable on a machine (cf. Blass and Gorevitch 2001). This results in a computer which can be used as a testbed for linguistic theories in order to show that a linguistic theory actually works.

It is the architecture of RRG that raises the idea that RRG is a linguistic theory which can be implemented with ease as it already has a design, and which can be translated into a procedurally executable code and, therefore, can be formalized without problems. However, as shown in Gottschalk (2012a; in press), several problems occur if one attempts to directly translate the linking algorithm into either a pseudo-code meta-language or into machine code. It is a particular step in the linking algorithm which is problematic. This step is given in (1)

below:

(1)

Step 2:

Determine the actor and undergoer assignments following the actor-undergoer hierarchy. (Van Valin 2005: 136)

In Gottschalk (2012a; in press), this step was formalized as a program snipped in pseudo-code:

(2)

```
algorithm step2
if number_argument_slots = = 1 in logical structure do
        if lexical_entry_verb = = takes_undergoer do
                undergoer = referring_expression_x;
        else
                actor = referring_expression_x;
        end if.
if number_argument_slots = = 2 in logical structure do
        actor = leftmost_argument;
        undergoer = rightmost_argument;
end if.
if number_argument_slots = = 3 in logical structure do
        actor = leftmost_argument:
        undergoer = new.choice( );
        non_macrorole = new.choice( );
end if.
```

If the number of argument slots in the logical structure equals 1, access to the lexicon is necessary to determine whether the lexical entry of the verb suggests that it can only be satisfied by an undergoer, or whether it can be satisfied by an actor. The default situation is when the verb takes an actor as the only macrorole in intransitive verbs with only one argument slot. There is only one situation where the AUH can apply as the sole basis for the determination of macroroles: when the verb is transitive with two argument slots in the

logical structure (cf. (2)). From an RRG perspective, being transitive is the default situation for verbs. This default situation results in many theory-internal considerations in which the AUH is an essential part of the theory, since the assignment of macroroles in the lexicon is the marked situation (cf. Van Valin 2005: 66). The crucial situation with respect to the computability of RRG where RRG runs into difficulties occurs with three argument slots in the LS. RRG cannot account for three-place predicates solely based on its procedural approach and the AUH (cf. (2)). In cases with variable undergoer linking, as in English three-place predicates, the linking algorithm cannot determine to which of the possible candidates for the undergoer the corresponding macrorole should be assigned (cf. Gottschalk in press).

(3)

a. [**do**´(Starbuck, Ø)] CAUSE [PROC & INGR **have**´(Apollo, security key)]

b. Starbuck [Actor] gave the security key [Undergoer] to Apollo

c. Starbuck [Actor] gave Apollo [Undergoer] the security key

Based on Van Valin's (2007) AUH, it is not possible to decide which argument should be assigned as undergoer if the computer is supposed to generate a construction as in (3a) or (3b). The reason for this is that the modified AUH leaves a choice between assigning the lowest ranking argument in the LS as undergoer, as in (3c), or assigning the second lowest argument in the LS as undergoer, as in (3b). Van Valin does not give an indication of how to solve the problem of variable undergoer linking in RRG (cf. Gottschalk in press). This is the reason for the use of the function 'choice' in the pseudo-code. All the function states is that it is the speaker's choice to assign undergoer to one of the arguments in the LS. Van Valin (2007) deals with this situation in that he uses preposition assignment rules. These rules cannot apply to this part of the algorithm, however, since, following Van Valin (2005: 136), the assignment of prepositions takes place in step 3 of the linking algorithm. Additionally, the AUH developed in Van Valin (2007) leaves a choice with respect to the assignment of three-place predicates.

A possible solution to solving the problem of RRG not being computationally adequate is presented in Gottschalk (2012a; 2014). Here, a lexicalist approach to RRG is chosen which does not use semantic macroroles in any way. This approach rather uses a unification-based approach to RRG by employing inheritance networks and attribute value matrices to populate logical structures in the mental lexicon. It is possible to extend this approach and develop a

computationally adequate model of RRG which employs conceptual graphs [CGs] (Sowa 2000) instead of RRG's Dowty-based LSs. Based on a procedure developed in Petersen (2007), it is possible to develop a computationally adequate RRG linking algorithm which is lexically motivated and can be implemented in a semantic parser.

The point of departure for using conceptual graphs to represent RRG semantics for Winther-Nielsen (ms) is that it is possible to create a formal language which is accessible for non-linguist learners. The reason why I choose conceptual graphs as a possible semantic meta-language in RRG is based on considerations which come from theoretical computational linguistics and have to do with the computability of RRG as discussed earlier in this section - conceptual graphs are computable on a computational device and they can apply to languages like Biblical Hebrew and are hence useful for a semantic parser of Biblical Hebrew.

Conceptual graphs are directed bipartite graphs. They have two kinds of nodes: Concepts and Relations. While concepts can stand alone, relations always are connected with at least one concept (cf. Petersen 2007; Sowa 2000). Formally conceptual graphs are equivalent with directed acyclic graphs and attribute value matrices, and they can be generated from them. Concepts in conceptual graphs consist of a type and a referent. Types can either be taken from an ontology, or they can consist of an individual to which the concept refers. The description of concepts can either be as simple as an existential quantifier, or they can be indexical. In the first case, the referent is simply left blank as a form of short hand and, in the latter case, it points to a specific referent by established conventions (cf. Petersen 2003). Relations are nodes modifying one or more concept nodes. They can bring two or more concept nodes into relation with each other (cf. Petersen 2003). The relations for RRG suggested in Gottschalk (2012a) are AGNT (agent) and PTNT (patient), which refer to Fillmorian grammatical relations. In my approach to RRG, these relations have replaced traditional semantic macroroles as suggested by Van Valin (2005). In conceptual graphs, a relation has a relation type, which determines its signature. Such a signature dictates which types the concepts attached to a relation can have (cf. Petersen 2003). Both relations and concepts are semantically defined by an ontology.

When it comes to the current semantic representation in terms of logical structures as they are used in Van Valin (2005), Nolan (2004) has shown that they can be represented as a unification grammar in terms of attribute value matrices [AVM]. Two examples of Nolan's

AVMs are given in figure 2 and figure 3 below.

(2)

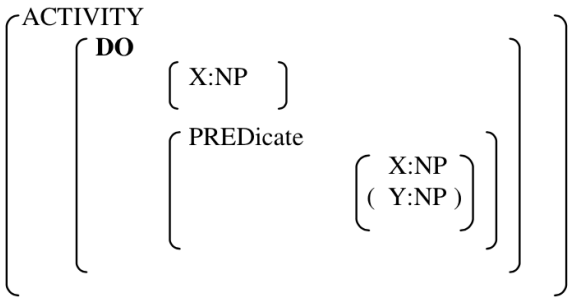

Figure 2 AVM for Activities

(3)

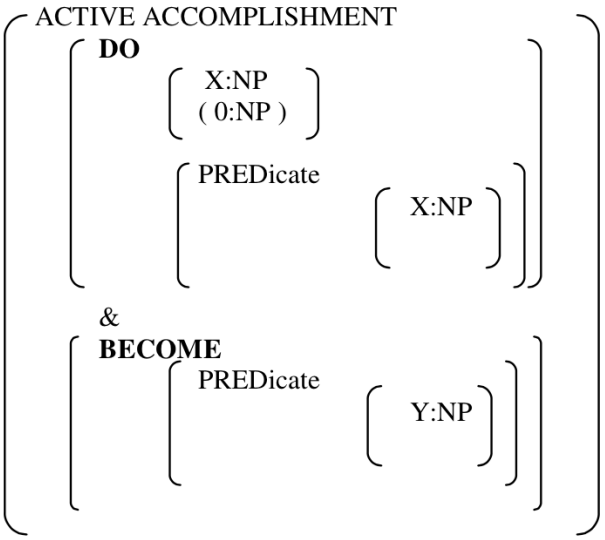

Figure 3 AVM for Active Accomplishments

Formally, AVMs are equivalent to acyclic graphs. Generally, when it comes to the computability of graphs, typically three possible variants are found regarding their complexity: a) those which are practically computational b) those which belong to context free grammars and related kinds of grammars with cubic computing time c) those where computability is not possible.

The AVMs representing the semantic structure in RRG can be expressed by the following formal grammar given in (4):

(4)

Non-terminals
State
Activity
Achievement
Accomplishment
Active Accomplishment
DO
Pred
INGR

Terminals
X
Y
Z

Production Rules

Aktionsart -> State
Aktionsart -> Activity
Aktionsart –> Achievement
Aktionsart -> Accomplishment
Aktionsart -> Active Accomplishment

Activity	->	Do
Do	->	X PRED
PRED	->	X Y

Active Accomplishment	->	Do Become
Do	->	X PRED
PRED	->	X Y
BECOME	->	PRED
PRED	->	Y

Accomplishment	->	Become
BECOME	->	PRED
PRED	->	Y

Starting Symbol
Aktionsart

In the example in (4), I have focused specifically on a grammar which generates AVMs for Activities, Active Accomplishments and Accomplishments. The grammar clearly shows that the data structure, which formally describes Van Valin's LSs, AVMs, can be generated by a context-free grammar. Does that mean that the whole theory of RRG is context-free and hence computable? The answer is: no, not necessarily. It is necessary that certain parts of the grammar need to be expressed by a context-sensitive grammar. This is, of course, compatible

with the model described here, but it causes RRG to have a less optimal complexity within the Chomsky hierarchy.

Although it is the case that the approach to semantics Van Valin uses in RRG is computable on a machine it is nevertheless the case that it is not suitable for languages like Biblical Hebrew for the reasons given in section 2. This is why I am developing a new approach to semantics in RRG. In the course of doing so, I will also developed a computable apporach to the syntax-to-semantics linking algorithm, which as presented by Van Valin (2005) is not a fully-fledged algorithm. Rather it is a guiding principle based on which an algorithm can be generated, and this is what I will be outlining in the next paragraphs.

A first step on the road to the development of this new approach to the syntax-to-semantics linking which should be implemented within semantic parser I seek to develop is the architecture of the mental lexicon, which is at the heart of this lexicalist approach to RRG. It is given in figure 4 below:

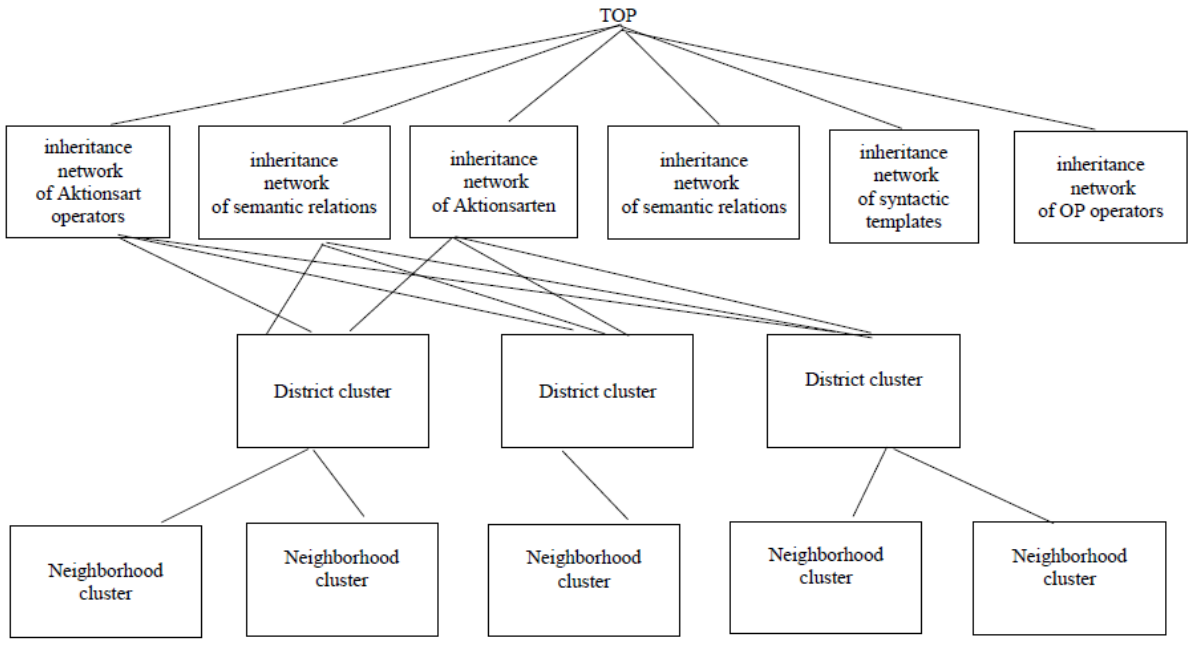

Figure 4 Architecture of the RRG lexicon

The mental lexicon is organized as an ontology. This will make it possible to generate conceptual graphs with it. The ontology defining types for the conceptual graphs is stored in the mental lexicon. It is an extension of the various ontologies developed in Gottschalk (2010)

and Gottschalk (2012a; 2014). In this ontology, not only the nodes from the LSC are stored and semantically defined by attribute value matrices, but also the ontologies for thematic relations, which replace macroroles in my approach, and the Aktionsart operators are stored in it. The ontologies for thematic relations, LSC operators, Aktionsart operators and Aktionsarten themselves make up the upper ontology, which is displayed in figure 1. Deeper in the ontology, semantic clusters in which verbs, nouns and other parts of speech are stored by their semantic meaning are stored in *district clusters* and *neighborhood clusters.* A district cluster contains words with less semantic content (in the sense of being less specific), like *go* in comparison to *amble*, while verbs as specific as *amble* are stored in neighborhood clusters.

With respect to computational complexity, the choice of an ontology which has the data structure of a tree has the advantage that trees are searchable with specific search algorithms in a reasonable time so that access to a lexicon which is structured in this way supports the attempt to develop a computationally adequate model of RRG.

In this model, the number of positive and negative features in the AVM defining them defines specific meaning of verbs. An AVM for the verb *receive* is given in figure 2 below:

receive
< > == [- transfer from x argument in LS]
<selectional properties> = = transfer
<manner of transfer> = = transfer of a non-abstract entity realized as y argument to x argument. Transfer of a non. abstract entity realized as z from y argument to x argument. Describes a neutral way of transfer to x argument.
<thematic relation> = = *give-rel*
<pointer to construction repository> = = ^[RP_{Agent} V [RP | PN] Recipient RP_{Theme}]
^[RP_{Agent} V RP_{Theme} [PREPPN | RP]$_{Recipient}$]
^[RP_{Agent} V RPTheme [PREP indef de N]$_{Recpient}$]

Figure 5 Lexical Entry for the verb *receive*

In a concrete implementation of the vocabulary learning tool, it could be possible to use WordNet as a reference ontology for the district clusters and neighborhood clusters which are suggested in Gottschalk (2010; 2012a) and to, in this way, generate glosses with the exact meaning of a Hebrew word. This is essentially also what Winther-Nielsen (ms) suggests.

All together, these clusters form the ontologies for the conceptual graphs, which constitute the semantic representations in this approach to RRG. In the lexicon, which is crucial to this version of a computational version of RRG, underspecification and AVM feature structures are employed to represent word meanings in the lexicon and templates for the LSC which are stored in the syntactic inventory, and which is an extension of the lexicon (cf. Nolan 2004; Van Valin 2005). Nolan (2004) has shown that it is possible to generate an AVM representing the LSC in RRG from the ontology in figure 4. The AVM for the LSC is given in figure 3 below and it is based on the immediate dominance rules developed in Van Valin and LaPolla (1997). These rules are given in (5).

(5) Immediate dominance rules

SENTENCE	→	{(DP)} CLAUSE
DP	→	XP /ADV
CLAUSE	→	{(ECS)}, CORE, (PERIPHERY), {NP*}
ECS	→	XP /ADV
PERIPHERY	→	XP / ADV
CORE	→	ARG*, NUC
NUC	→	PRED
PRED	→	V /XP
ARG	→	PRO / NP / PP

Figure 6 gives an example of an AVM for an LSC:

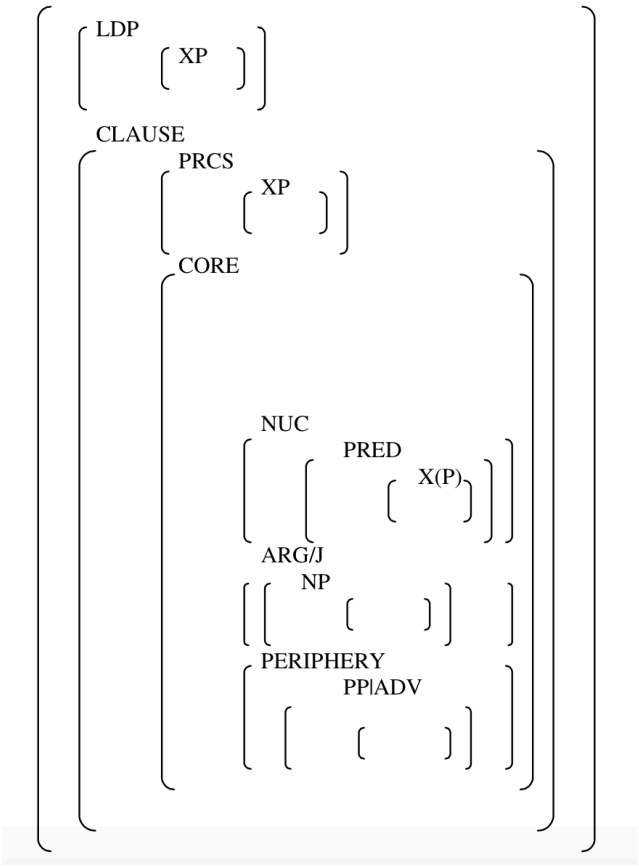

Figure 6 AVM representing the LSC

A similar approach is suggested for the operator projection. Van Valin and LaPolla (1997) have developed ID / LP rules to represent the operator projection. The rules are given in (5).

(6)

SENTENCE	→	CLAUSE	←	IF
CLAUSE ← IF	→	CLAUSE	←	OP*
CLAUSE ← OP	→	CORE (← OP*)		
CORE (← OP*)	→	NUC (← OP*)		
NUC (← OP *)	→	V/XP		

(6) is an example of a context-sensitive part of RRG, and it raises the question of how this part can be improved. For now, I will leave it to future research to investigate this question. A representation of the operator projection as an AVM is given in figure 8 below:

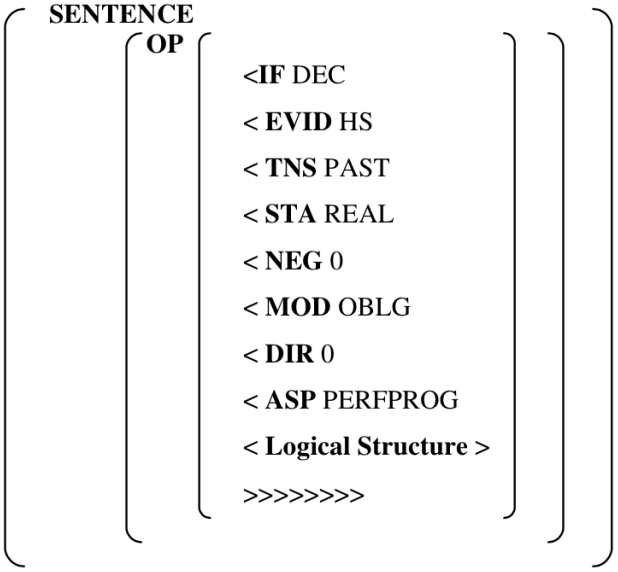

Figure 7 AVM representing the Operator Projection

For English, Van Valin (2005) suggests a set of syntactic templates to generate more complex syntactic templates from it. He assumes that there is a natural ordering for them within the syntactic inventory as part of the lexicon. It is possible to assume a subsumption order based on their structure and their order (cf. Nolan 2004: 214). Instead of Van Valin's approach, I use AVMs similar to the one in figure 2 in a hierarchical order, which are then attached to the ontology in figure 1. As in that ontology, more specific AVMs, which are a subset of another AVM, are stored deeper in the ontology because they are attested as being suitable in computational implementations of RRG, as shown in Nolan (2004).

The conceptual graphs for the semantic representation of RRG are formally equivalent to the AVMs representing RRG semantics suggested in Nolan (2004). The advantage of using conceptual graphs, as already mentioned, is that they are more general data structures for representing the semantics of a text compared to those in FunGramKB, as they also apply to ancient languages like Biblical Hebrew. Furthermore, they are already attested as computationally derivable with the help of a computational device. An approach to generating them via a computer has been put forward in Petersen (2007). I am employing Petersen's approach to generating them to a limited degree to establish a computational approach to RRG. The conceptual graphs presented in this approach are similar to the pedagogical ones presented in Winther-Nielsen (ms). Conceptual graphs for all RRG Aktionsarten are given in 7 below:

(7)

a. [Person: Mulder]<-(Expr)<-[**do'**: drive]->(Pat) → [vehicle: car]
activity

b. [Person: Scully] → (Agent) → [**INGR:** win] → (Thm) →[reward: FBI-medall]
achievement

c. [Person: Director Skinner] → (Expr) → [SEML: cough]
semelfactive

d. [Substance: Kryptonite] → (Expr) → [PROC: glim]
process

e. [Substance: Kryptonite] → (Pat) → [PROC & INGR: melt]
accomplishment

f. [Person: Mulder] → (Expr) ← [**do'** & INRG**:** drive] → (Loc) → [Destination: Area 51]
active accomplishment

The formal representation of the conceptual graphs in (7) uses types for concepts, which are stored in the lexicon within the neighborhood clusters and the district clusters. Concepts in the graphs in (7) are represented in square brackets, while relations are represented in parentheses. In the examples in (7), types are *person, vehicle, reward substance* and *destination*. The relations in the parentheses contain the thematic relations defined in Gottschalk (2012a). Aktionsart operators such as **do´**, INGR, SEML, PROC, PROC & INGR, and **do´**& INGR are stored in the ontology defining the hierarchical structure in the lexicon.

Sowa (1992) explains that the ISO standard of conceptual graphs completely expresses the full semantics of common logic. In common logic, quantifiers can range over functions and relations but common logic retains a first-order style of model theory and proof theory. In order to support a higher-order syntax without the computational complexity of higher-order semantics, the CL model uses a single domain D that includes individuals, functions, and relations

Sowa (1992) gives a formal grammar for conceptual graphs within Extended Backus Naur Form rules. Since the Extended Backus Naur Form is equivalent to context-free grammars and rather is a different from of expressing them, in (8) I have developed a context-free grammar generating the new conceptual graphs for RRG:

(8)

Conceptual Graph	-->	{concept	conceptual relation}*
Concept	-->	concept type	referent
Conceptual relation	-->	relation type	
Relation type	-->	signature	

What the grammar rules for the new approach to semantics in RRG show is that conceptual graphs of the type in (8) can be generated by context free grammars as in (9), and so are not less optimal than Van Valin's (2005) approach to semantic structures in RRG. The advantage of using conceptual graphs instead of a semantic meta-language as developed in FunGramKB is that conceptual graphs can be used to create a machine-readable lexicon for languages like Biblical Hebrew, and not only for modern languages like English and Spanish.

A flow chart of the syntax-to-semantics linking algorithm employing conceptual graphs rather than logical structures is given in figure 5. It is based on Petersen (2007):

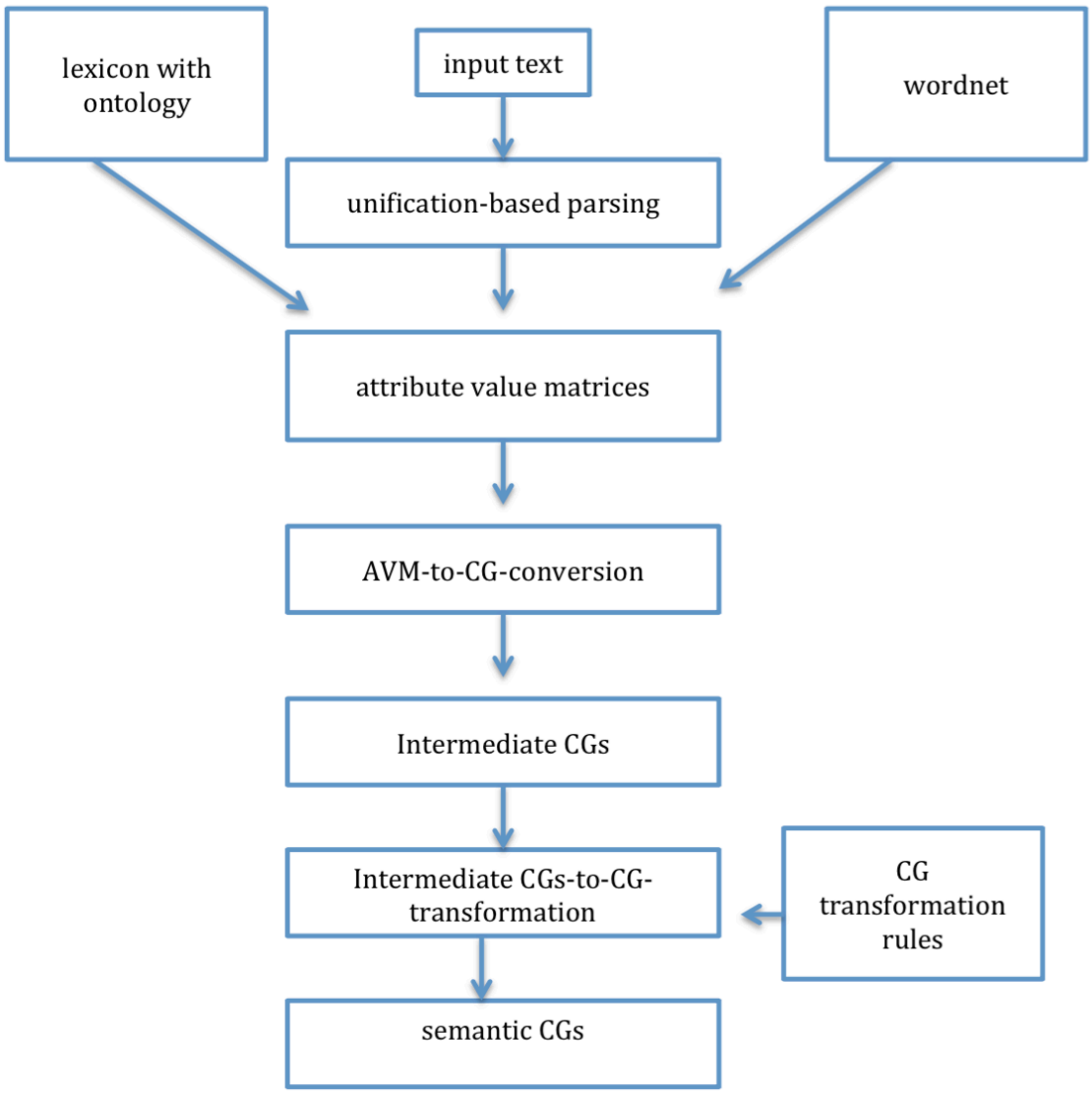

Figure 8 Workflow for the generation of semantic structures in RRG using conceptual graphs instead of logical structures developed by Van Valin (2005)

The AVMs representing the LSC are decomposed as CGs. This is done in two steps: in a first step by transformation rules, which match each node in the AVM representing the LSC against concepts in the ontology to generate an intermediate CG, and in a second step, these intermediate CGs generate a semantic CG. The transformation rules operate at the word level and at the phrase level. Here, the rules are looped over the nodes in the AVM to produce a fragment of a CG. Each node in the AVM is used as input to a rule, which has a specifically privileged concept called the attachment point. It is distinguished from all other nodes by having a unary relation called *attach*. This is the point where the CG fragment is joined to another CG fragment. These rules constitute a new linking algorithm, which is now a completely rule-based procedure which can be executed by a computational device. These rules are illustrated in tables 1 and 2 below:

Table 1 phrase structure rules

input	output
SENTENCE →{(DP)} CLAUSE	{(DP)} CLAUSE ← attach
DP → XP /ADV	XP / ADV ← attach →
CLAUSE →{(ECS)},CORE, (PERIPHERY), {RP*}	{(ECS) CORE ← attach (Periphery){NP*}
ECS → XP / ADV	XP / ADV ← attach
PERIPHERY → XP / ADV	XP / ADV ← attach
CORE → RP *, NUC	{RP ← attach} *, NUC ← attach

Table 2 word rules

Input	output
Noun, singular from operator projection	[*noun*] ← attach
Noun, plural from operator projection	[*noun*: {*}] ← attach
Verb, verbal operator from operator projection is not "be_X"	[*verb*] ← attach
Verb contains operators from lexicon and ontology, is "be_X"	attach → [state: [be]- attr → [*verb*]]

When it comes to the semantic parsing of Biblical Herbew, it is the output ¶[*verb*] <-- attach where Winther-Nielsen's algorithm for the determination of Aktionsarten in Biblical Hebrew comes into play. The algorithm is represented in pseudo-code in (10) below:

(9)

if

Causative: Causation **C** + CAUS → CAU

elseif

Process (y):	Accumulated result	U	+ PROC	+TEL	→ ACC
elseif **Incident** (x):	Instantaneous result	U	+ INST	+TEL	→ ACH
elseif **Activity** (x,y):	Completed activity	A	ACTV	+ TEL	→ ActACC
elseif **Semelfactive** (x):	Instant activity	A	+ INST	–TEL	→ SEML

elseif

Activity (x (y)):	Ongoing activity	A	ACTV		→ACT
elseif **State** (x,(y)):	Unchanged situation	U	UNCH		→ STA

end

then assign verb-specific semantic roles: is the x A or U and is there a y?

(Winther-Nielsen ms)

The input to the revised algorithm for the determination of Aktionsarten for Biblical Hebrew is stored in the WordNet Ontology so that, based on that information, it is possible to derive the correct meaning of words like נתן nātan within their morpho-syntactic context and generate a gloss with the exact meaning of נתן nātan in the glosses in Bible Online Learner. Just as in the approach in Gottschalk (2012a; 2014), Winther-Nielsen does not use Van Valin's (2005) semantic marcoroles Actor and Undergoer, but rather uses a number of lexical semantic relations, which he derives based on the valence of the verb. This way, it is possible to point to constructional schemas as necessary in the way it is described in Gottschalk (2012a; 2014). Winther-Nielsen's algorithm for the determination of the semantic role comes into play in the last step in table 2. It is given in (11) below:

(10)

if **Single argument states** → x (U)

11	State or condition:		PATIENT
12	Existence:		ENTITY

elseif **Two argument non-verbal states** → x, y

26	Possession:	POSSESSOR,	POSSESSED
01	Attributive:	ATTRIBUTANT,	ATTRIBUTE
02	Identificational:	IDENTIFIED,	IDENTITY
03	Specificational:	VARIABLE,	VALUE
04	Equational:	x, y=REFERENT	

elseif **Two argument verbal states** → x (A), y (U)

21	Pure location:	LOCATION,	THEME
22	Perception:	PERCEIVER,	STIMULUS
23	Cognition:	COGNIZER,	CONTENT
24	Desire:	WANTER,	DESIRE
25	Propositional attitude:	JUDGER,	JUDGEMENT
27	Internal experience:	EXPERIENCER,	SENSATION
28	Emotion:	EMOTER	TARGET

elseif **31-35 Single argument activities** → x (A)

31	Unspecified activity:	EFFECTOR
32	Motion:	MOVER
33	Static motion:	STATIC-MOVER

34 Light emission: L-EMITTER
35 Sound emission: S-EMIITER

elseif **41-45 One or two argument activities** → x (A),y (U)

41	Performance:	PERFORMER	PERFORMANCE
42	Consumption:	CONSUMER	CONSUMED
43	Creation:	CREATOR,	CREATION
44	Directed perception:	OBSERVER	STIMULUS
45	Use:	USER	IMPLEMENT

end

The intermediate CGs generated by the rules in tables 1 and 2 generate an intermediate CG of the type: *[SENTENCE [CLAUSE [CORE [RP: Mulder], [NUC: sees], [RP: Scully]]]*]. Based on the lexicon and CG-transformation rules, a conceptual graph as given in 6 is generated from the intermediate CG. With the two sets of rules table 1, it is completely possible to replace the semantics-to-syntax linking algorithm as developed in Van Valin (2005) with a computationally tractable set of rules to develop a computational processing model of RRG.

All these changes in the framework of RRG result in a new architecture for RRG, which will be realized in the semantic parser, which uses RRG as its linguistic engine. The new architecture of RRG is given in figure 6: In this approach to RRG, the lexicon is divided into different parts and consists of a construction repository and a syntactic repository. These two repositories are formally defined by the lexical ontology presented in 1. Complex inheritance processes within the lexicon feed the linking algorithm with the necessary information and make a link from semantics to syntax and from syntax to semantics possible.

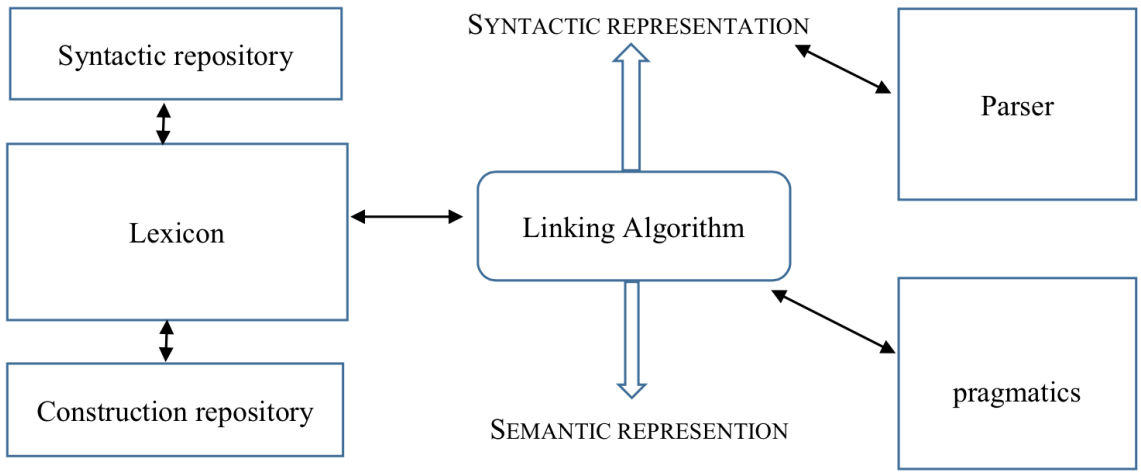

Figure 9 The new architecture of RRG which uses the lexicon to a greater extend and which adds a construction repository to the theory

5 Conclusion

The paper has presented a novel way to determine the meaning of verbs for learners using Bible Online Learner. With the theoretical model presented, it will be possible to develop a semantic parser based on a new approach to the syntax-to-semantics linking algorithm in RRG, which will improve corpus-driven vocabulary learning by generating the exact meaning of a specific group of verbs in Biblical Hebrew in the glosses displayed in Bible Online Learner. Furthermore, it was, on the one hand, shown that parts of RRG already exist which are computationally optimal, namely the generation of LSs, but which are not suitable for ancient languages like Biblical Hebrew.

Based on these findings, it was possible to develop an approach to RRG which is a computationally adequate model of the syntax-to-semantics linking algorithm which does also apply to languages like Biblical Hebrew. This solution works based on a WordNet ontology which Petersen (2007) enriched with conceptual graphs and which is extended into a computational model of RRG. This will no doubt make it possible to come up with a semantic approach to RRG in Biblical Hebrew studies, and will improve corpus-driven computer-supported language learning in general.

Let the Bible be your tutor – A case study on contextualized corpus-driven vocabulary learning with Bible Online Learner and Learning Journey

Judith Gottschalk
Aalborg University, Rendsburggade 14, 9000 Aalborg
gottschalk.judith@gmail.com

Abstract

How can computer-supported vocabulary learning be improved by employing corpus-driven and contextualized language learning? This paper suggests enabling learners to learn the vocabulary of Biblical Hebrew within the morphosyntactic context the words occur in as an approach for better vocabulary learning. The idea is to use an algorithm for the automatic determination of Aktionsarten for Biblical Hebrew developed by Winther-Nielsen (ms) to enable users of Bible Online Learner and Learning Journey Online to derive lexical rules from a corpus. Additionally, the role of detailed learning statistics and gamification is discussed in this paper, and an approach to self-directed vocabulary learning is developed.

1 Introduction

The acquisition of second language vocabulary is a difficult issue. Students use various approaches to acquire words in a foreign language. In language classes and learning groups, students communicate in the target language. Some students read texts in the foreign language and look up every word they do not know in a dictionary, while others watch movies in the target language, or even write letters or stories in the language which is to be acquired.

In his dissertation on acquiring Biblical Hebrew vocabulary, Thompson (ms) discusses a number of learning strategies. Learning strategies are mental, physical or creative aids for learning a subject, e.g. vocabulary, more successfully, and with regards to language learning, there exists a vast plethora of strategies which support different learning types in

remembering and actively using acquired vocabulary. One of the learning strategies can roughly be described as 'us[ing] a computer program to practice words', and, indeed, this strategy seems to be effective in the case of learning second language vocabulary, as tools like Anki (http://ankisrs.net) or Memrise (http://www.memrise.com) promise.

Anki is a tool that provides its users with digitally editable flash cards. Its users can create decks with several hundreds digital learning cards, which they can use to memorize facts and vocabulary. In the course of a learning session, the user clicks through a deck and decides how long it should take until a specific card is shown again. If the user knows a word very well, the time until he is shown the card next can be extended. Anki uses the SuperMemo Algorithm, which I will describe in some detail in section 2. In Memrise, on the other hand, a user can define quizzes for himself. Thus it is, for example, possible to ask 'What does the verb נתן nātan' mean, and, then, up to four possible translations can be given. The user can click on the solution which he thinks is right, and is then immediately shown the correct answer. Memrise uses the metaphor of gardening, and uses gamification to motivate the user to stay with the tool and keep learning. The learning approach with Memrise will likewise be discussed in section 2.

Nevertheless, it has been pointed out several times that learning words using methods such as word lists and flashcards as in Anki and to Memrise means that words are encountered outside of the authentic context in which they occur and this causes problems when it comes to the active as well as the passive use of acquired vocabulary (cf. Nagy 1997; Thompson 2011). Thompson cites three studies showing that computer-supported language learning offers good opportunities for a hybrid approach combining learning vocabulary lists and contextualized learning. While Cobb (1999) and Groot (2000) focus more on advanced learners, Kang (1995) suggests that computer-supported vocabulary learning using contextualization is a promising approach. In his approach, Kang has tested the effects of using the computer to introduce words within their genuine context, and was able to show that presenting words along with both written and visual contexts was much more successful for definition recall, listening comprehension and knowledge transfer than paired-associate learning with paper and pencil (cf. Kang 1995; Thompson ms).

Software such as Anki and Memrise demonstrate that learners do access tools for computer-supported vocabulary learning nowadays, and, even earlier, there has been use of computer-

supported methods to support vocabulary learning. Nevertheless, it is rather rare that words are learned through contextualization with a glossed corpus. Thompson (ms) gives an overview of some studies and I paraphrase his literature review here, but what becomes obvious is that this literature is rather outdated. Chun and Plass (1996) found that students who read a text on a computer screen are able to learn the vocabulary if they have the possibility of looking up the words easily, or if the words in the text are glossed with both text and pictures (cf. Thompson ms). The success of this method is that it leads to both good short-term and long-term retention. Following Lyman-Hager (2000), glosses enhance the comprehension of a reader, especially if the text contains a high incidence of unknown words. Leffa (1992), on the other hand, has investigated the use of hypertext technology using an electronic glossary for the acquisition of reading comprehension. This study has shown that an electronic glossary is more efficient than a traditional dictionary. The electronic dictionary allows the subject to understand 38% more of a passage in less than 50% of the time (cf. Leffa 1992; Thompson ms). Thompson (ms) sums these findings up and explains that students using computers for reading begin reading much earlier and begin to better gain the aspects of vocabulary knowledge which are otherwise acquired through implicit learning.

Recent research in the area of vocabulary learning has shown that learning vocabulary in context, i.e. providing learners with vocabulary which is relevant to the context in which the learner currently acts, can enhance the learning experience (cf. Dearman and Troung 2012). This is a different kind of context than that which the aforementioned literature has focused on, and it shows that it is necessary to define clearly whether *context* means *the context in which a word occurs in within a text* or *a context to which a word belongs based on the location where the learner is situated*. A concrete example for such a situation is when a learner uses a cell phone with an app for vocabulary learning, and the cell phone, based on GPS data, recognizes that the learner is currently in a café, and therefore sends a notification to the learner with a vocabulary question related to 'cafés' (cf. Dearman and Troung 2012). This is one possibility for using contextualization and it is quite similar to the approach in Memrise: by creating a mental bridge between some known thing, e.g. the café one is currently situated in, and the vocabulary which needs to be learned, a mnemonic is created which supports learners in memorizing the vocabulary in question. This technique is widly used in computer-supported language learning, and a number of contextualized language learning tools have been developed in recent years which provide learners with prompts and content that is relevant to the objects they are currently interacting with (cf. Baudin et al.

2007; cf. Ogata and Yano 2004), or a situation they are currently part of (cf. Al-Kekhlafi, Hu and Zheng 2009).

What this discussion indicates is that learning vocabulary in context is a wide area of application, and tools like those developed by Dearman and Troung (2012) usually apply to modern languages like English, Danish and German, but not to ancient languages like Biblical Hebrew. The reason is that, for languages like Biblical Hebrew, typical contexts are restricted to the Hebrew Bible and contexts from this source do not usually apply to modern situations like visits to cafés or cinemas, or other places where people spend time nowadays.

Nevertheless, despite the issue described above, Bible Online Learner is a program for computer-supported language learning of an ancient language which *does* use contextualization. The tool is corpus-driven. By providing learners with a fully-glossed corpus of the Hebrew Bible from the Eep Talstra Center for Bible and Computer [ETCBC], learners are provided with a state of the art resource with which they can explore the Hebrew Bible. The vocabulary that the learners acquire is learned in context, namely the context of the Old Testament and the Hebrew Bible. In Bible Online Learner, the learners can solve quizzes and complete grammar drills, and they can practice vocabulary based on word frequency.

In spite of these features, Bible Online Learner indeed has an Achilles' heel. While Bible Online Learner uses contextualization to help the learners acquire their vocabulary as they learn their words by reading the Hebrew Bible (like in Anki and Memrise), exercises employed for vocabulary learning in Bible Online Learner do not support contextualization. Instead, learners are provided with lists of words derived from the corpus without context. Additionally, Bible Online Learner is too sensitive to typos which users make, and it does not recognize whether or not a learner has already mastered a word.

Winther-Nielsen (ms) and Gottschalk (accepted) both discuss the role of the Hebrew verb נתן nātan, which, depending on the morphosyntactic context the word appears in, has several different meanings. Clearly, for words like נתן nātan, it makes only very little sense to show the lexeme *without* its morphosyntactic context because even if Bible Online Learner accepts translations of נתן nātan based on its different meanings, learners are not able to derive lexical rules which enable them to learn what the word means in which context. The problem here is that the ETCBC corpus provides the learners with all possible translations of the word

in its glossings.

As a result of findings regarding difficulties learners face when they cannot properly recognize a word because determining meaning is problematic, Winther-Nielsen (ms) and Gottschalk (accepted) developed a formal framework within Role and Reference Grammar [RRG] using conceptual graphs to automatically derive the exact meaning for נתן nātan from the corpus. The approach in Gottschalk (accepted) is to use a semantic parser to correct the glossings in the corpus. This way, learners are only exposed to a correctly glossed corpus. Furthermore, Winther-Nielsen has developed a pedagogical algorithm using the context of Aktionsarten, which enables the learner to derive the meaning נתן nātan based on its morphosyntactic context. Gottschalk (accepted) suggests using this algorithm within Bible Online Learner to enable the learners to acquire lexical rules, which in turn helps them to determine the meaning of נתן nātan based on the context its tokens occur in. It will be the task of this paper to practically show how this idea from Gottschalk (accepted) can be used in a persuasive approach to vocabulary learning in Biblical Hebrew.

Bible Online Learner comes with a tool called Learning Journey, which enables teachers and facilitators to provide learners with feedback on their learning progress with Bible Online Learner. Learning Journey is a performance-optimizing tool which uses detailed and elaborated learning statistics for every learner using Bible Online Learner (cf. Gottschalk and Winther-Nielsen 2013). The system works as follows: while a learner is completing a quiz in Bible Online Learner, the time he spends on an exercise, the number of correct and incorrect answers he provides, and the answers he gives are recorded and submitted to the statistics database once the learner is finished with the quiz. Learning Journey analyses these training values and calculates various parameters from the data, which provides inside knowledge concerning whether the learner is successful with his learning or whether there are areas where he can improve. How the learning statistics in Bible Online Learner work compared to those in Anki and Memrise is described in detail in section 2.

The problem this paper deals with is as follows: How can vocabulary learning with Bible Online Learner and Learning Journey be improved, and how might the theoretical approaches in Winther-Nielsen (ms) and Gottschalk (accepted) apply to provide learners with a persuasive tool for language learning? These questions will be discussed in the remainder of this paper, which is organized as follows: In section 2 the background of this study is

presented. I discuss the SuperMemo algorithm used in Anki in this section, give an overview of how vocabulary learning with Memrise works and introduce corpus-driven language learning with Bible Online Learner and Learning Journey. Section 3 introduces my case study with students from the Fjellhaug International University College Denmark. In this section I analyze their responses to questions in a qualitative questionnaire in the light of grounded theory to develop a design approach for contextualized vocabulary learning. This design approach will be presented in section 4 together with suggestions for improving Bible Online Learner and Learning Journey in order to support a stronger involvement of learning statistics in contextualized vocabulary learning with the two tools.

2 Background

Since 1967, there has been one dominant approach in audio-supported language learning, and this is called the Pimsleur System. This system uses a series of 30-minute audio lessons in which a basic vocabulary and phrases are introduced and reviewed in a cued-recall fashion. The recall-fashion uses a graduated interval recall using a progressive series of exponentially expanding intervals both within and across the lessons (cf. Edge et al. 2012). This system clearly has the advantage that it uses deliberate overlearning to grow confidence in a learner's language skills and to enable him to immediately use the language. However, on the downside, within such lessons, the repetitions are not scheduled in real-time, there is no feedback, and the learner must make time for his lessons, additionally the lessons are limited in number, which means the learner can run out of lessons (cf. Edge et al. 2012).

The Leitner System was developed at roughly the same time as the Pimsleur System, and can be used for physical flashcards. It also uses spaced repetition by creating numbered flashcard piles where a subsequent pile represents flashcards of an increased memory length and increasing inter-study intervals (cf. Edge et al. 2012). The idea is that cards move from one pile with longer memory length to a pile with shortened memory length as soon as they are not remembered correctly. This way, cards which are more difficult are practiced at shorter intervals and, hence, more often. The problem, however, is that learners rarely know when to introduce new flashcards and, furthermore, the piles could easily become unmanageable (cf. Edge et al. 2012).

The SuperMemo Algorithm is a computer-assisted approach to managing and scheduling the

review of digital flashcards. It is used in tools like Anki. SuperMemo targets a fixed retention rate of the flashcards where this is worked out as "the optimum interval I_R following a repetition R of an item is calculated as the product of I_{R-1} and the "optimum factor" for items of that "easiness" at repetition R. The table of optimum factors is updated over time, in response to achieved recall rates, and with an aim of converging on a long-term recall rate of 95%" (cf. Edge et al. 2012: 433). SuperMemo has a self-assessment facility using a 6-point scale to update the easiness of an item to be learned. This algorithm has the advantage of adaptive spaced repetition which evolves according to the performance a learner exhibits over time. However, the disadvantage is that it does not support initial learning, with the first test following item introduction not typically occurring for at least 5 days (cf. Edge et al. 2012). Edge et al. characterizes the use of the SuperMemo algorithm in Anki as follows:

> Anki has three distinguishing features: cards support multiple directions of cued recall between multi-attribute "facts" (e.g., the meaning, pronunciation, and appearance of a Chinese character); tests are scheduled each day at a fixed learner-specified rate; and after being tested on a card the learner selects one of four options indicating when they want to be tested on it next. These all increase learner load in terms of deciding which attributes to test, keeping up with scheduled sessions, and judging recall performance. (Edge et al. 2012: 422).

Memrise, on the other hand, employs three scientific principles: elaborate encoding, choreographed testing, and scheduled reminders. The idea of elaborate encoding is that what a learner already knows is related to what he seeks to learn. One approach is to use taxonomic categories and to organize words into common groups, allowing the learners to see connections between them and to use cues helping them to later recall the vocabulary (cf. http://www.memrise.com/science/#elaborate-encoding). Choreographed testing uses an approach also mentioned in Metsämuuronen (2012): regular testing results in better recognition. This is close to the idea of using something like the SuperMemo Algorithm in Anki (cf. http://www.memrise.com/science/#choreographed-testing). Finally, the idea of scheduled reminders is that users of Memrise receive notifications and reminders that they should proceed with their learning (cf. http://www.memrise.com/science/#choreographed-testing).

The approach to vocabulary learning in Bible Online Learner and Learning Journey differs to a great extent to the strategies employed in Anki and Memrise. Bible Online Learner provides learners with software that can display the appropriate meaning of a word in its syntactic context, but the downside of this approach, however, is that the students are not forced to use

their knowledge of grammar to derive the appropriate meaning of a word in the syntactic context and, hence, they rarely memorize the correct translation of a word (cf. Winther-Nielsen ms). Extensive evaluation by Winther-Nielsen (2013a) during a course in Copenhagen has shown that better means for learning vocabulary with Bible Online learner is necessary. The current solution in Bible Online Learner is that Hebrew glosses are displayed with their frequency of occurrence in the Hebrew Bible. How this is done for the Hebrew verb נתן *nātan* is shown in figure 1.

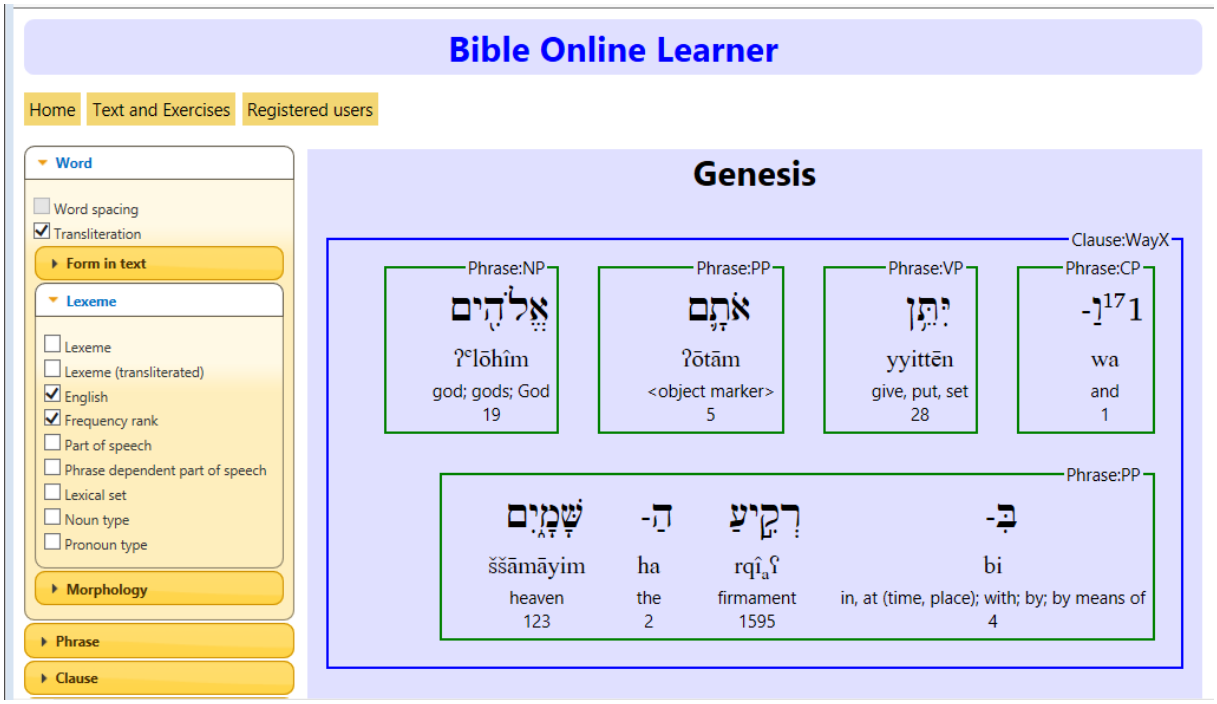

Figure 1 Display of glosses and frequency in Bible Online Learner

The idea was that learners could explore the glosses in the corpus and practice memorization of the meaning of a word in the context of a clause. In case of נתן *nātan*, three glosses are displayed: *give, put,* and *set*. The verb is ranked at position 28 in the list of relative frequency for all words in the Hebrew Bible. In Bible Online Learner, in turn, the learners are provided with exercises based on the frequency of the words. The various lexemes are listed, and the learners can input a free translation of the word into a textbox. Figure 2 below gives an impression of how vocabulary learning with Bible Online Learner works. Here, words with a high frequency like nātan can be practiced in a similar way just as is done with tools like Anki.

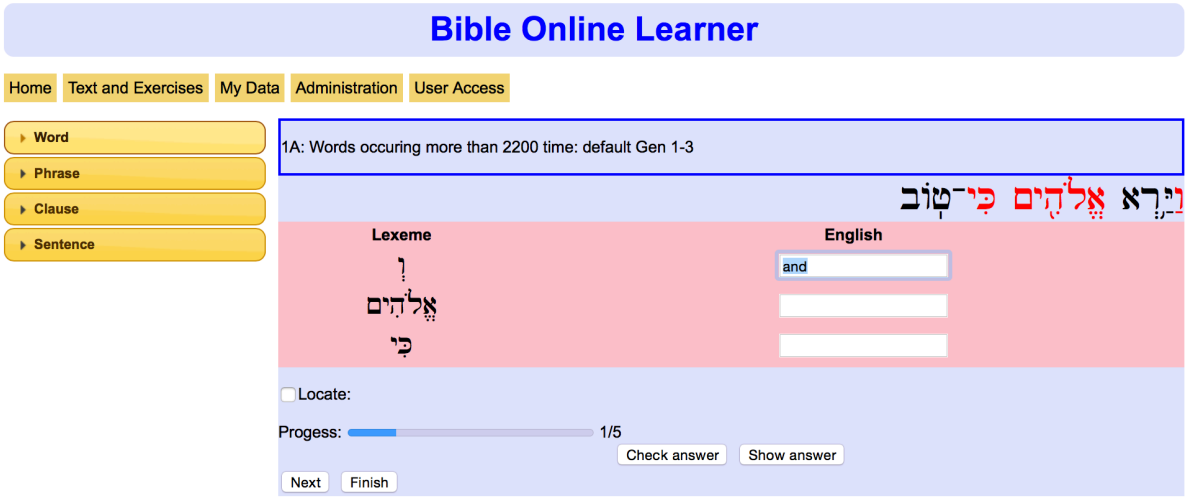

Figure 2 Vocabulary exercises in Bible Online Learner containing words with a frequency of more then 2200 times in the Hebrew Bible

Data on how fast a student was able to solve a quiz in Bible Online Learner (and whether her answer was correct, and what kind of mistakes she made etc.) is automatically transferred to the server on which Learning Journey runs. The tool is a performance-optimizing tool that is used by teachers to monitor learners while they learn Hebrew. Teachers see how their students progressed in their learning, and they can encourage them and point them to areas with room for improvement. In the future, we plan a role and a permission model for Learning Journey, where the students can also keep track of their learning themselves.

With Item Response Theory it is possible to statistically measure the performance of a learner in order to predict the probability of her response by establishing the position of the individual learner along a line of some latent dimension. The use of IRT is quite prominent in educational environments and, in this context, the latent trait is often called *ability*. Moreover, it is possible to measure how difficult to acquire an item which is to be learned actually is using IRT (cf. Parchev 2004: 5).

The application of IRT within systems of language learning ranges from personalized systems for Computer-Assisted Language Learning [CALL], which recommends appropriate English vocabulary according to the statistically evaluated difficulty level a learner can master (cf. Chen and Chung 2007: 624), to systems which use fuzzy IRT to support personalized e-learning based on student profiles (cf. Chen and Duh 2008).

While the SuperMemo algorithm in Anki and similar tools is based in the factor of the learners' self-estimation in combination of a means to calculate the best time to repeat an item, Learning Journey uses precise measures from psycholinguistics to determine if a learner as acquired a learning item as well as measures from IRT. The tool uses measures like *accuracy* and *proficiency* and *ability*, which are rooted in psycholinguistics and statistics research. In the formal framework I developed together with Nicolai Winther-Nielsen and which is introduced in Gottschalk and Winther-Nielsen (2013) and Gottschalk (ms), the concept of automatization is also used, which is discussed in DeKeyser and Criado (2012). The parameter *proficiency* is defined as in (1):

(1) Proficiency = sum of right answers / right answers per minute

Winther-Nielsen and Hery (2013) have shown that a low response time is not a perfect indicator for successful learning. The reason for this is that some students can achieve a high speed in answering the quiz questions while at the same time submitting a very high number of wrong answers. This is the reason why Winther-Nielsen and Hery suggest calculating proficiency as in (1) (cf. Gottschalk ms). *Accuracy* is calculated as in (2) below:

(2) accuracy = (sum of right answers + sum of wrong answers) / sum of wrong answers

What this parameter shows is the degree to which learners have responded correctly compared to the degree of wrong answers (Gottschalk ms). In section 4 I will show how the ability of Learning Journey to exactly measure the learning progress of a student can be integrated in to vocabulary learning in Bible Online Learner. What I would like to point out here is that using measures which are rooted in the actual learning performance of a student rather than in their ability to answer a learning item correctly as in the SuperMemo Algorithm is a much more elaborated way to determine if a student has mastered to learn a word as has been pointed out above based Winther-Nielsen's and Hery's (2013) work.

In his most recent work, Winther-Nielsen explores the application of Role and Reference Grammar [RRG] (cf. Van Valin 2005; cf. Gottschalk 2010; Gottschalk 2014) in improving vocabulary learning with Bible Online Learner, and expands the theory of RRG with a solution using a WordNet ontology (http://wordnet.princeton.edu/), and following a proposal made by Petersen (2007). His idea is to enrich parser output to generate conceptual graphs

[CG] from it (cf. Winther-Nielsen ms). By employing Petersen's (2007) approach, Winther-Nielsen develops a pedagogical framework which enables learners of Biblical Hebrew to derive the meaning of a word in a specific syntactic context with the help of a simple algorithm which uses the theoretical basis of RRG.

Gottschalk (accepted) is based on the work by Winther-Nielsen, and adds the theoretical basis for extending Bible Online Learner with a semantic parser to improve the automatic generation of glosses in the tool. The idea is simple: since verbs like נתן nātan in Hebrew have multiple meanings depending on the morphosyntactic context the verbs occur in, and since, as shown in figure נתן, the ETCBC corpus used in Bible Online Learner provides learners with all possible translations for a word like nātan, learners do not have the possibility of learning by exploration of the corpus. They are not able to derive lexical rules to know when which meaning of nātan occurs because the corpus provides them with all *possible* translations rather than with the one which fits the context. However, in his paper, Winther-Nielsen (ms) has shown that there are means of determining when nātan needs to be translated as *give, put* or *set.* On the one hand, the semantic parser, which Gottschalk (accepted) describes, can be used to correct the ETCBC corpus. Nevertheless, Gottschalk (accepted) and Winther-Nielsen (ms) only give a rough idea of how their theoretical approach to improving vocabulary learning with Bible Online Learner can be applied practically in the tools. Here are the key questions addressed in this paper: how can computer-supported vocabulary learning with Bible Online Learner and Learning Journey be improved regarding a more successful method of contextualization which meets the requirements of an ancient language like Biblical Hebrew, and how can the theoretical findings from Gottschalk (accepted) and Winther-Nielsen (ms) be applied practically within the tools?

3 Empirical study of vocabulary learning with Bible Online Learner

The design approach chosen in this study is user-centered design (cf. Sanders 2005). I have been in contract with learners using Bible Online Learner who have filled in an anonymous qualitative questionnaire.

I am using grounded theory to analyze the qualitative data to develop a scientific theory from the systematic analysis and interpretation of the empirical data at hand (cf. Rogers et al. 2011; Gottschalk 2012b). Glaser and Strauss (1967) were the first to develop this approach. When

using grounded theory, the task is to develop a theory that fits the data which has been collected, and a grounded theory is developed by alternating data collection and the analysis of this data (cf. Rogers et al. 2011; Gottschalk 2012b). Usually, the first step is to collect qualitative data via questionnaires or focus group interviews, or in the case of this study by directly getting into a discussion with the prospective users of a piece of software, and to make them co-create the tool by giving detailed input. In step two of the application of grounded theory, further data collection is conducted and the data are interactively analyzed. This step continues until a proper theory is developed and new insights into a specific topic are obtained (cf. Rogers et al. 2011; Gottschalk 2012b).

I carried out a study with nine users of Bible Online Learner to find out how vocabulary learning with the tool worked for them. I collected qualitative data. The results of this study in a line-by-line analysis of participants' experiences are given in tables (1) to (4)[1] below:

(1)

What was your strategy for learning the Hebrew Vocabulary?	
Answer	**Line-by-line analysis**
Paradigm table. Flash Card. Recite.	1. Classical Learning Strategies (paradigm table, flash cards), 2. Using contextualized and decontextualized learning.
Iphone app	1. Using a computer as learning strategy
A large mixture : Bible Online Learner, books, Private teacher, Memrise, listening to lessons.	1. Contextualized learning (Bible Online Learner) 2. De-contextualized learning (Memrise) 3. Other strategies (teacher, listening to lessons)
I printed out the word list and did them chronologically; I wrote a "wing" when I had them learned (sic.). I used a bit memrise to make it more fun now and then.	1. Decontextualized learning without a computer 2. Decontextualized learning with a computer 3. Desire for fun
Memrise.com and text reading. Mainly Memrise.	1. Decontextualized vocabulary learning with a computer
To refresh Hebrew learned long ago.	1. No learning strategy identifiable
The use of the word list from Bible Online Learner used in Memrise.com while using Paradigms Master Pro.	1. Decontextualized vocabulary learning with a computer

[1] In the tables below I am directly reporting comments from non-native speakers and the comments have been reproduced exactly, language errors included, to preserve their authenticity.

I have mostly practiced with my homemade flashcards.	1. Decontextualized vocabulary learning without a computer
The Memrise app.	1. Decontextualized vocabulary learning with a computer

(2)

What caused you the most problems learning vocabulary?	
Answer	**Line-by-line analysis**
Verbs and phrase functions.	1. Problems with the morphology of verbs and other phrases
I didn't have problems because I have used my app (sic.).	1. Computer-supported vocabulary learning is successful.
The test was different from Bible Online Learner which shows words in a context	1. The test at DBI did not fit the contextualized teaching method offered by corpus-driven language learning with Bible Online Learner
Keeping up, set of the time to it (sic.).	1. Problems in the personal organization of the learning.
Not being attentive to differences between Ayin and Aleph, Waw and Vet, and so on.	1. Problems with spellings which are very similar.
Just needed to experience how the IT system "thinks"!	1. It was not clear whether Bible Online Learner works.
Words being very similar.	1. Problems with developing lexical rules.
Motivation.	1. Missing the 'fun factor' in learning.
The ones alike[2]	1. Problems with developing lexical rules.

(3)

If you did not use Bible Online Learner for vocabulary learning, what was your reason for this?	
Answer	**Line-by-line anaylsis**
Confused. Need clarification from Hebrew books.	1. Problem with the corpus-driven approach of exploring the corpus by reading from the Bible.
I havent (sic.).[3]	1. No data available

[2] The participant is referring to words which are alike.
[3] The participant is referring to the fact that he did not use Bible Online Learner.

I used it mostly to read and translate text.	1. Using Bible Online Learner for active exploration of the corpus and learning vocabulary in context. 2. Not the decontextualized exercises from Bible Online Learner.
Memrise is fore (sic.) [for]efficient and fun. In BOL you don't get credit if you can't remember whether it is a "," or a ";" if more meanings in the same word. Also you don't get arround (sic.) all the words, and you can sit praticing (sic.) the same words 3-5 times in row.	1. Missing gameful elements in Bible Online Learner like a reward system. 2. Problems with decontextualized vocabulary learning with the tool 3. Problems with monotony due to repetitions in vocabulary learning.
Lot of the same words again and again.	1. Problems with monotony due to repetitions in the vocabulary learning.
I used Bible Online Learner.	1. Using Bible Online Learner for contextualized and decontextualized vocabulary learning.
If you could remember more synonyms of a word you had to use varying punctuation.	1. Problems with the decontextualized exercises in Bible Online Learner.
To see if I could recognize the words in the biblical context.	1. Bible Online Learner is used as tool for corpus-driven language learning.

(4)

How do you think vocabulary learning with Bible Online Learner can be improved?	
Answer	**Line-by-line analysis**
Included (sic.) vocabulary in pdf. for download. Revise when not on internet or computer.	1. Enabling decontextualized vocabulary learning via generated lists.
I dont (sic.) know how it works, I cant (sic.) even find where the vocabulary is.	1. No familiarity with the tool.
If it can include audio. Some (sic.) to read the words/ sentences.[4]	1. Need for an additional channel to learn the vocabulary.
One should be able to know, that you practice all words within frequency rank (XXX-XXX),. fx 345-374. This skill, Memrise lacks. And again; no difference between "," and ";" when typing (it also gives an (sic.) inaccurate test-information). Not having to face the same words so often,	1. Clearer organization of the exercises. 2. Bug report. 3. Reduction of repetitions in the tool. 4. Interaction between Bible Online Learner and Learning Journey.

[4] The pariticipant suggests that Bible Online Learner could include audio recordings, which she / he can play on her /his computer.

and if Bible Online Learner could follow your progress, and stop showing words you already do master (sic.)	
It works very well. Just a few remarks like, the verb is the basic form/Infinitive, the object written like: <object marker>	1. Necessity to improve the glossings in Bible Online Learner.
More gamification, streamlining, better word translation.	1. Need for gamification. 2. Improvement of glosses.
I haven't used it that much so I really don't know.	1. No data available.
?	1. No data available

Based on the findings in a line by line analysis of the learner's responses, it is possible to derive the axial coding, which is displayed in figure 3 below:

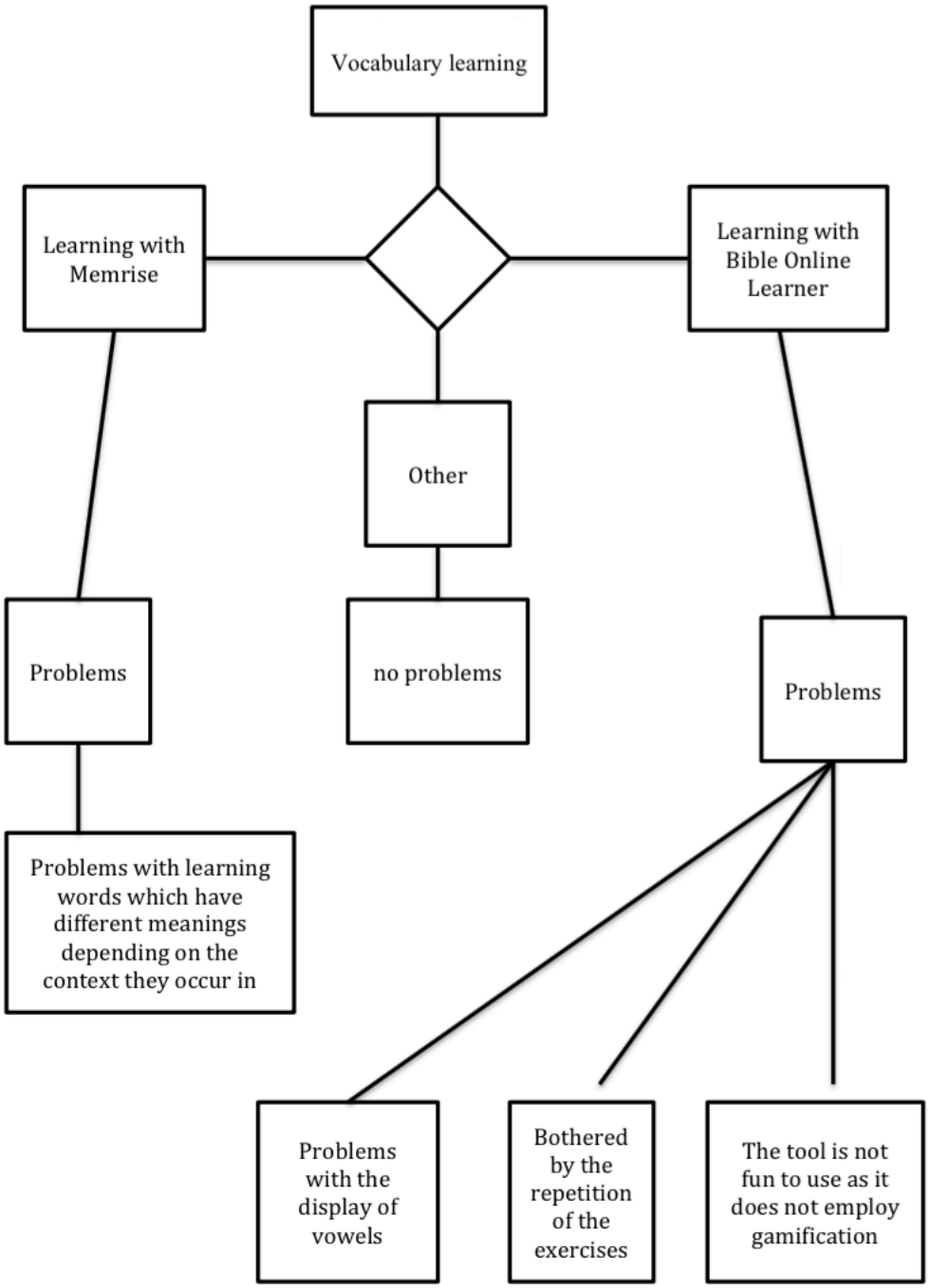

Figure 3 Axial Coding for the development of a grounded theory for computer-supported vocabulary learning

Like in a kind of activity diagram in unified modeling language [UML], the main points which can gained from the line-by-line analysis in tables 1 to table 4 are displayed in figure 3. These show that the students have chosen four broad learning strategies, and most of those strategies were supported by a computer. Participants used either only Memrise, Bible Online,

or other methods. Most of the students used only Memrise to learn vocabulary, and here the most problems occurred (which are put together in figure 3 under the heading as 'Problems with learning words which have different meanings depending on the context they occur in) The main problem for the students was that, with Memrise, they acquired the Hebrew words out of context and, since many words look similar, have similar meanings or more than one translation, they got confused due to the fact that, in Memrise, vocabulary is acquired in decontextualized quizzes rather than in the context of the Old Testament as in Bible Online Learner. When it comes to the use of Bible Online Learner, the students mostly report that two things caused them problems: handling the different signs for vowels in the tool, that the tool did not motivate them in the same way as Memrise as it does not use gamification and that it repeated the same exercises too often. When the students used other methods, relatively few problems occurred.

Nicolai Winther-Nielsen taught the learners who filled in my questionnaire at the Fjellhaug International University College in Copenhagen. He shared his personal analysis on the semester he spent teaching the class with me (his evaluation is added to this case study in the appendix). Following Winther-Nielsen the problem was that in the first two weeks of the semester he has spent too much time on lecturing on verb stems instead of focusing on an important midterm test. This test covered the most important vocabulary of Biblical Hebrew. At the same time the students started using Memrise for vocabulary learning to pass the test and did not use the videos on grammatical phenomena Winther-Nielsen has provided them with and skipped class.

The class was scared of the midterm test in advance and 50% of the class failed in the test which resulted in more students focusing on using Memrise for the preparation for the final exam and less students used Bible Online Learner. Also many students stopped going to class and they did not prepare for class, which made teaching even more difficult. In a second test in October again two good students failed in the test. One failed because it was hard for him to learn with the glosses in Bible Online Learner and one because he did not have sufficient time for practicing Hebrew. A third student has general problems to learn Hebrew and failed in the exam. A third test is passed by an online student and another student fails again in November. Finally in December only two gifted students of Winther-Nielsen are able to follow the crucial training on Hebrew syntax in class. The final exam took place on December 18th and still needs to be corrected.

In his evaluation of the course Winther-Nielsen comes to the conclusion that most of the students have not made enough time for practicing Hebrew. His course is developed for a daily workload of 5 - 6 hours a day and most of the students in his class have not been able to meet that goal. On the other hand the students prefer enforcement and punishment in their classes rather than persuasion; the students in Copenhagen have not been motivated enough to practice for so many hours as self-directed learners and rather returned their responsibility to learn as self-directed learners back to the teacher from whom they expect to make them feel bad if they to do not perform well. Possibly the punishment the students asked for could have been resolved if Bible Online Learner had used gamificaion to make the learning more fun.

Nevertheless Winther-Nielsen thinks the most crucial point, which caused the high fail-rate in the course was the decontextualized vocabulary learning with Memrise. This can also been seen in my analysis in the figure 3 and tables 1 to 4. Using decontextualized vocabulary learning caused the students to not be able to derive lexical rules to determine the meaning of words like נתן nātan in their morphosyntactic context and hence prevent them to learn the vocabulary properly.

Another group of students in Madagascar taught be Christian Højgaard had more success. In his class 39 out of 49 students passed the exam. Højgaard's class consisted of two groups of students: A minor group consisted of students knowing English and being familiar with computers which meant that they could easily use Bible Online Learner and also used it a lot. The major group had troubles using Bible Online Learner because they were not familiar with neither computers nor the tool. In general, the students using Bible Online Learner had far better results than the students not using the tool. Generally however the students did not use decontextualized vocabulary learning with Memrise which can give an inside in why the students in Madagascar performed better then the students in Copenhagen.

These findings support Winther-Nielsen's (ms) and Gottschalk's (accepted) hypothesis that providing the students with an ideal corpus of the Hebrew Bible and enabling them to acquire lexical rules to be able to derive the meaning of the Hebrew vocabulary from the context is a fruitful approach. This is essentially the case because, as displayed in figure 3, learning decontextualized vocabulary with Memrise mainly causes the problem that students are not

able to differentiate words which are either similar in their meanings, or similar in some other way.

The students made various suggestions for how learning with Bible Online Learner can be improved. Most of the students who used Bible Online Learner complained that words are repeated too often, and that the system does not recognize when they have mastered a word. Additionally, the system is not as fun to use as Memrise because it lacks gamification, and, finally, Bible Online Learner is not as easy to use as originally intended when the tool was developed.

From these findings it is possible to derive the following theory about computer-supported language learning: Learners using tools for computer-supported vocabulary learning benefit most from their learning if they have a) the possibility to learn vocabulary by exploring a hypertexed corpus with unambiguous glossing which perfectly match the meaning of a Hebrew word. Additionally rather than learning vocabulary with decontextualized quizzes (as in Memrise) learners benefit from learning vocabulary in contextualized way, i.e. if they learn words within their morphosyntactic context. Exercises should be adapted to the learning progress a learner makes while using the tool and the tool should be fun to use for the learners.

With this grounded theory in mind it will be possible to design a new means for vocabulary learning in Bible Online Learner and Learning Journey.

4 Concrete design approach

Bible Online Learner and Learning Journey are both applications which use PHP on the server-side, and JavaScript on the client-side, as well as a MySQL database. They are reachable over the Internet at http://bibleol.3bmoodle.dk and http://statdb.3bmoodle.dk, or they can be installed on any server with a Linux operating system, an Apache Webserver, MySQL and PHP within a local network, as is done at SALT in Madagascar. Users are registered by an administrator (usually a facilitator), and can use this account for logging on to both Bible Online Learner and Learning Journey. Bible Online Learner can also be accessed via a Google Login.

Once a learner has logged on to Bible Online Learner, she can either explore the ETCBC corpus with its glosses, or she can solve exercises to train a skill, as described in section 2. Learning Journey, on the other hand, provides its users with a web interface where facilitators and learners can explore their learning progress in terms of learning statistics. Learning Journey uses various data mining algorithms, and the formal IRT-based framework used by Metsämuuronen (2013). A facilitator is, for example, able to evaluate the learning progress of his student and motivate the learner by giving him feedback based on the learning data, and point him to things he should practice more deeply, or where he should focus more in his grammatical drills. The learning statistics are so detailed that the facilitator even gets an overview of what types of mistakes a student makes. On the other hand, it is also possible to predict learning progress with the tool, by using IRT (cf. Gottschalk and Winther-Nielsen 2013).

The analysis of the user feedback and Nicolai Winther-Nielsen's analysis in section 3 and the encoding of the line-by-line analyses in tables 1to 4 in figure 3 indicate that what caused the students problems with learning vocabulary was that Memrise, while being fun and easy to use, does not provide the learners with a context in which the vocabulary occurs. With regards to the verb נתן *nātan*, Winther-Nielsen illustrates the challenge for the learners as follows:

> To illustrate the challenge for the learner, the verb נתן *nātan* is used 5 times in Genesis 1-3. For the example from Gen 1:17 in Fig. 1, the learner has to deduce from the context that the luminaries are the referents for the masculine plural clitic pronoun attached as suffix to the object marker את֫ *ʔet*. To 'give' them 'in the firmament' clearly does not involve an animate receiver, and therefore the two synonymous glosses 'put' and 'set' are more appropriate for placing the luminaries in visible view in the sky. In Genesis 1:29-30 נתן *nātan* has the meaning 'give' and is followed by a receiver 'to you' in a prepositional pronom (sic.) followed by the gift itself, 'all herbs and trees'. (Winther-Nielsen ms. 3)

When being faced with such a word in a tool like Memrise, learners definitely experience trouble identifying the meaning of nātan without the morphosyntactic context the word occurs in. Of course, the learners can search for the translations in a dictionary and choose the one which might fit the rest of their translation. However, this way of hunting for dictionary interpretations and learner attempts to match a preferred Bible translation from memory prevents students from learning the lexical rules which they can memorize and automate for language acquisition.

The problem is that neither the approach in Memrise nor the current approach to exercises for vocabulary learning in Bible Online Learner use contextualization for vocabulary

acquisition. Based on this, Gottschalk (accepted) makes the following suggestion for improvements: instead of providing the learners with multiple lexemes outside of their syntactic context as is currently done in Memrise and Bible Online learner and asking them to enter all possible translations in a textbox, the learners will be provided with text excerpts from verses in the Bible with the lexeme occurring in its syntactic context.

Winther-Nielsen (ms) has developed an algorithm which makes it possible to derive the meaning of a verb from its semantic context, and learners are 'forced' to follow this algorithm before they can enter the meaning of a verb into a textbox. Winther-Nielsen's algorithm is given in (3) below:

(3)

if

Causative:	Causation	**C**	+ CAUS		→ CAU
elseif					
Process (y):	Accumulated result	U	+ PROC	+TEL	→ ACC
elseif **Incident** (x):	Instantaneous result	U	+ INST	+TEL	→ ACH
elseif **Activity** (x,y):	Completed activity	A	ACTV	+ TEL	→ ActACC
elseif **Semelfactive** (x):	Instant activity	A	+ INST	–TEL	→ SEML
elseif					
Activity (x (y)):	Ongoing activity	A	ACTV		→ACT
elseif **State** (x,(y)):	Unchanged situation	U	UNCH		→ STA

end

then assign verb-specific semantic roles: is the x A or U and is there a y?

if **Single argument states** → x (U)

11	State or condition:	PATIENT
12	Existence:	ENTITY

elseif **Two argument non-verbal states** → x, y

26	Possession:	POSSESSOR,	POSSESSED
01	Attributive:	ATTRIBUTANT,	ATTRIBUTE
02	Identificational:	IDENTIFIED,	IDENTITY
03	Specificational:	VARIABLE,	VALUE
04	Equational:	x, y=REFERENT	

elseif **Two argument verbal states** → x (A), y (U)

21	Pure location:	LOCATION,	THEME
22	Perception:	PERCEIVER,	STIMULUS
23	Cognition:	COGNIZER,	CONTENT
24	Desire:	WANTER,	DESIRE
25	Propositional attitude:	JUDGER,	JUDGEMENT
27	Internal experience:	EXPERIENCER,	SENSATION
28	Emotion:	EMOTER	TARGET

elseif **31-35 Single argument activities** → x (A)

31	Unspecified activity:	EFFECTOR
32	Motion:	MOVER
33	Static motion:	STATIC-MOVER

34 Light emission: L-EMITTER
35 Sound emission: S-EMITTER

elseif **41-45 One or two arguments activities** → x (A),y (U)

41 Performance: PERFORMER PERFORMANCE
42 Consumption: CONSUMER CONSUMED
43 Creation: CREATOR, CREATION
44 Directed perception: OBSERVER STIMULUS
45 Use: USER IMPLEMENT

end

With this algorithm, it is possible to determine the Aktionsart for a verb like נתן nātan, which in turn enables the learner to derive the meaning of the word in its morphosyntactic context. The idea for improving exercises in Bible Online Learner, therefore, is that the words which the users should learn are highlighted within their syntactic context in an excerpt from the Bible, and before entering the meaning of the verb in the textbox, the learners are asked specific questions which help them to derive the meaning of the word such as 'Is the Aktionsart of the verb a causative?'. Here, the user can choose a radio button with 'Yes' or 'No', and is directed to the next question he has to answer: 'Does the verb have the Aktionsart process?'. Again, the learner has to click a radio button and is, this way, directed to the correct translation and forced to follow a clearly defined algorithm to derive the meaning of a word so that he will be able to develop knowledge of lexical rules for Biblical Hebrew vocabulary. Figure 4 shows a sketch of a possible interface within Bible Online Learner for this kind of exercise:

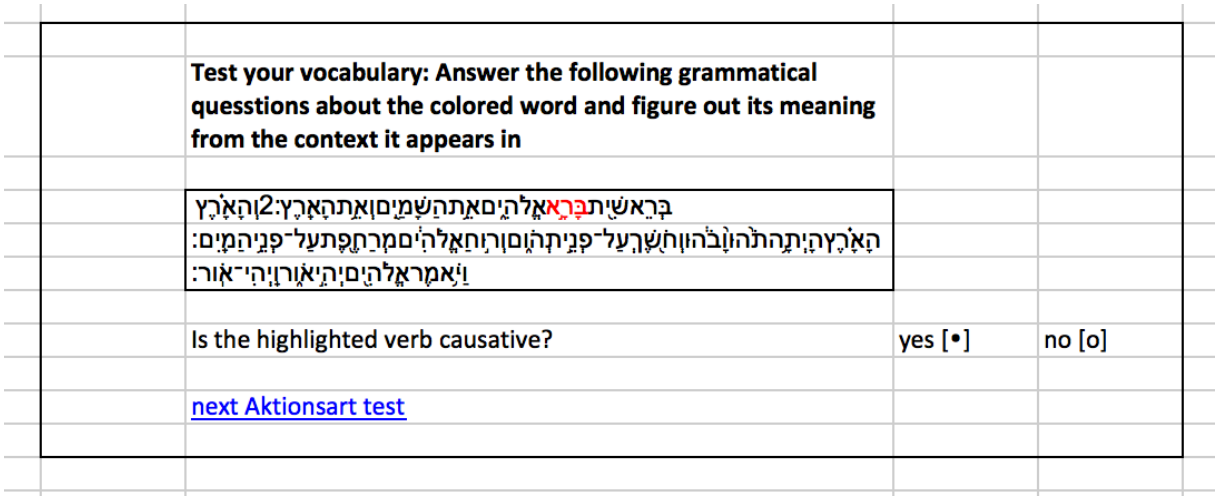

Figure 4 Example for an interface for contextualized vocabulary learning with Bible Online Learner and an RRG-based Aktionart test

By using this kind of interface, and by repeatedly learning how to apply the RRG-based

algorithm developed by Winther-Nielsen (ms), learners are supported in constructing rules as part of the process of language acquisition. This is also where the RRG-based semantic parser, which Gottschalk (accepted) has developed the theoretical basis for, comes into play. Instead of providing the learners with several translations in the glosses in Bible Online Learner, the task will be to automatically derive the translation which exactly fits the syntactic context a verb like נתן *nātan* occurs in, that in the case of the gloss in Gen 1:17 *put* is a better translation than *give*. To scaffold the realization that in addition to providing the learners with contextualized exercises, the glosses in the ETCBC corpus will be revised with the implemented semantic parser developed in Gottschalk (accepted) to provide the learners with only thoroughly checked examples when they use Bible Online Learner to explore Biblical Hebrew.

Another question the learners raised in their questionnaire answers is why Bible Online Learner repeats exercises for words which the learners have already properly mastered? A possible solution for the problem that Bible Online Learner does not *know* when a learner has mastered a vocabulary item is the following rough design for a better interaction of Bible Online Learner and Learning Journey was mentioned by Claus Tøndering in personal communication with the author and can be implemented by the author in turn of the further development of Learning Journey: In Bible Online Learner a frequency rank is contained which determines the frequency of each word in the corpus. This frequency rank is associated with each word in the corpus. Based on this information it is necessary to set exercises of progressing difficulty in Bible Online Learner in a chain. Here it is assumed that words with high frequency are easier to learn and words with lower frequency are moderate to learn whole rare words are difficult to lean.

With Learning Journey it is possible to determine if as student has mastered a difficulty level of the exercises. One possible approach could be to say, that an exercise with a set of easy words in Bible Online Learner is mastered if Learning Journey has calculated that the student has reached a specific percentage of the ability, accuracy and proficiency of the maximum values the learner could have reached potentially. This could then be put into relation in with a specific threshold the learner needs to meet in order to answer the next vocabulary question correctly. If that is the case the learner can move the moderately difficult exercises. Here in turn the student again needs a specific performance calculated by Learning Journey to move the final and most difficult exercises in the exercise chain. This way the advantages of the

detailed statistics which can be calculated with Learning Journey over an algorithm like SuperMemo can be practically used to reduce repeating exercises in Bible Online Learner and to better meet the students needs.

In order to implement such a solution the following things need to be implemented 1) A chaining mechanism which chains exercises with different difficulty levels with each other 2) A way that Learning Journey notifies Bible Online Learner that a student has reached the necessary percentage of knowledge to reach the next exercise level and 3) The teacher must design his exercises in a way that they can be chained. When I refer to exercises for vocabulary learning I mean specifically the kind of exercises as they are shown in figure 2 and designed in this thesis.

Opening Learning Journey up not only to the facilitators but also to the learners is a further way of improving vocabulary learning with the two tools, Bible Online Learner and Learning Journey. This way, the learners can use Learning Journey to monitor their learning statistics through self-monitoring and peers and teachers can keep track of the learning progress through surveillance.

The final remark the learners made in the questionnaire is that Bible Online Learner is not fun to use. The approach to solving this problem is to develop a Facebook-like friendship model which enables users to share their learning results with their peers and their teachers and to, this way, receive feedback from these groups by receiving comments or acknowledgements in the form of ‘likes’ or ‘faves’. If a student is friends with another user, he can share learning results with this user and make his learning progress public to this user. The student can decide whether or not to show his learning results to all students in the form of a broadcast, or whether he wants to show his learning results exclusively to one user, or to a specific group of users in the form of a multicast. Moreover, the student can decide if he would like to show his learning progress exclusively to the teacher of the course. This way, I will add a social aspect to the feedback already provided by the IRT-based learning statistics in Learning Journey. I believe with a combination of both learning statistics *and* social recognition in terms of collaboration via feedback, I can overcome the general flaws that an isolated social recognition function has (following Hamari and Koivisto (2013)). The student makes his learning progress public on a timeline (as on Facebook) which is to be found on the landing page of Learning Journey Online. This timeline can also be used by the teacher, course

creator or site admins to spread information to the users of the system.

In terms of visibility and privacy issues within the system, students can decide on their own if they would like to make their learning results public to the teacher and other students. The student has the following choices:

- Keep the data private and do not share it with anybody
- Make the learning data accessible to the teacher
- Make the data public to students which are friends but do not give the teacher access
- Make the learning data public to befriended students and the teacher
- Make the learning data public to everyone on the course

Based on findings in Gottschalk (2012) and Gottschalk (2013a), I believe that this technique supports the type of feedback which students desire following qualitative research, and that this specific game mechanism (which can also be found on various browser games on Facebook) increases user activity and, consequently, motivates learners to use the system to improve their language learning.

Winther-Nielsen (2013b) has developed a framework for persuasion in Bible Online Learner which is based on Fogg (2003). It differentiates three levels in the development of ability and motivation in learners through computer-assisted language learning. The first and simplest method for enhancing ability is reduction, which is well known from quizzes. It satisfies the basic needs of learners for reviewing knowledge and memorization. More sophisticated is tunneling, which proceeds in a predefined learning progression. Tailing is the most persuasive activation, where training with the persuasive technology is adjusted to the learner's knowledge level, age, learning style, progression, goals, and other highly individual parameters related to vocational needs (cf. Gottschalk and Winther-Nielsen 2013; Winther-Nielsen 2013a).

As it is at the moment, Learning Journey enables tunneling in learning, as it gives the teacher the possibility of adjusting his teaching to all students and providing the learners with predefined exercises. With the chained exercises and a closer interaction of Bible Online Learner and Learning Journey described in this section, tailoring is enabled and by using feedback via learning statistics, a feedback loop is initiated within Learning Journey and, in

this way, a pedagogical approach (as suggested by Laurillard (2012)) which is centered on constant communication between the entities in a learning process is employed (cf. Gottschalk and Winther-Nielsen 2013), which additionally enables motivation. This is the case as Hamari and Koivisto (2013) have shown empirically that social motivations and social influence (particularly in gamification elements involving reciprocal benefits) are strong predictors that systems using gamification increase user activity and are fun to use. Hamari and Koivisto (2013) have shown empirically that social motivations when related to social influence, specifically, when the users find reciprocal benefits from using gamification are strong predictors that a system using gamification has increasing user activity. However, the amount of recognition users receive during their use of the system does not directly affect their attitude towards gamification to a significant degree. Just as receiving 'likes' on Facebook is not enough, rather the user, as a consequence, must at the same time feel that receiving and giving recognition to other users using the system increases the benefits acquired from the system (cf. Hamari and Koivisto 2013).

This means that social recognition, when employed in the right way, is one way of giving feedback and to employ gamification in Bible Online Learner and Learning Journey to make the tools more fun to use.

5 Conclusion

In the present study, I have proposed a way to practically apply the highly theoretical approaches to vocabulary learning with Bible Online Learner developed in Winther-Nielsen (ms) and Gottschalk (accepted; this thesis) to improve vocabulary learning in tools for computer-assisted language learning.

Based on an empirical study, it was possible to pinpoint problems in vocabulary learning with tools for computer-supported language learning. Learning vocabulary out of the morphosyntactic context a word occurs in (as in tools like Anki, Memrise, and in the current exercise design in Bible Online Learner) causes learners problems with learning how to derive lexical rules to determine the exact meaning of a verb like nātan. When, however, vocabulary learning is done using a corpus-driven tool like Bible Online Learner, the learners can acquire vocabulary in context, which is a successful approach. A further way to improve

vocabulary learning is opening a performance optimizing tool like Learning Journey to support learners with detailed learning statistics and to empower learners to keep track of their own learning rather than using an automatized algorithm to determine if a learner has acquired a word..

It will be necessary to further develop Bible Online Learner and Learning Journey, however. In a first step, the semantic parser, which is described in Gottschalk (accepted; this thesis), needs to be implemented to clean up the glossings in the ETCBC corpus, and, in a second step, it will be necessary to merge Bible Online Learner and Learning Journey. This way, Bible Online Learner will be able to interactively react to input the tool receives from learners practicing vocabulary with the tool so that persuasive vocabulary learning as envisioned by Winther-Nielsen (ms) will be possible in the future.

General thesis conclusion

I started this master's thesis with Nicolai Winther-Nielsen's opening words of his class in at FIUC: It is not a teacher who teaches a learner Biblical Hebrew, but the Bible itself, when learning the language with Bible Online Learner and Learning Journey. In the course of an extensive introduction, where I gave an inside in how teaching Biblical Hebrew with Bible Online Learner and Learning Journey works and more generally what approaches to Hebrew didactics and what vocabulary learning strategies exist, I approached first to a highly theoretical approach to vocabulary learning in Biblical Hebrew using RRG and then I showed the practical aspects of how making the Bible teach students Hebrew can be more successful.

The idea in Gottschalk (accepted) which is part of this master's thesis is that computer-supported vocabulary learning using RRG as linguistic engine can be considerably benefit from using conceptual graphs (Sowa 2000) for the semantic description of a language. The reason is, while approaches in FunGram developed by Periñán-Pascual (2013) are successful for modern languages the same approach cannot work for an ancient language like Biblical Hebrew. (cf. Winther-Nielsen ms). If one seeks to improve corpus-driven language learning it is necessary to have a semantic parser for Biblical which uses a semantics framework which can deal with the semantics of an ancient language like Biblical Hebrew. Additionally the implementation of a semantic parser of course benefits from a theory of RRG which is tractable for *all* languages and not only for modern ones like Danish, English, German or Spanish.

When it comes to concrete approaches in vocabulary learning decontextualized vocabulary learning with Memrise is proved to be a less successful way to learn words in Biblical Hebrew than an approach where the Hebrew vocabulary is learned in context as in the new approach for vocabulary learning presented in this thesis. This was shown in the second paper, which is part of this master's thesis. The design of contextualized vocabulary learning with Bible Online Learner and the creation of exercise chains and a means for Bible Online Learner to communicate with Learning Journey to notify Bible Online Learner if a learner has mastered to learn a word will be a future task to implement. To implement exercise chains a considerable amount of design effort is required while the programming of this chain is relatively straight-forward. The biggest problem is most certainly a psychological one as this way of implementation requires the teacher to strictly stick to specific rules of designing

exercises. Of additional benefit for vocabulary learning over tools like Memrise and Anki beside contextualized vocabulary learning is the idea of using the sound framework of learning statistics provided by Learning Journey to determine if a learner has mastered to learn a new word rather than using the SuperMemo Algorithm employed in Anki as this algorithm does not use measures developed in psycholinguistic research.

With its detailed analysis of what makes vocabulary learning with Bible Online Learner and Learning Journey this master's thesis is the starting point for a future improvement of vocabulary learning. In a next step it will be necessary to develop an implementable and fine-grained design for both the semantic parser suggested in the first paper of this thesis and for the enablement of Learning Journey to notify Bible Online Learner if a learner has mastered the next exercise level in the exercise chain contained in Bible Online Learner.

References

Baudin, J.S., Intille, S.S., Tapia, E.M., Rockinson, R. and Morris, M.E. Context-sensitive microlearning of foreign language vocabulary on a mobile device. In *Proc. AmI 2007*, Springer-Verlag (2007): 55-72.

Behringer, Reinhold and Sinclair, Georgina. 2013. *IWEPLET 2013 Proceedings.* EuroPLOT Retrieved December 08, 2013 from http://www.lulu.com/shop/reinhold-behringer-and-georgina-sinclair/iweplet-2013-proceedings/paperback/product-21189131.html.

Carpenter, Bob. 1991. The Generative Power of Categorial Grammars and Head-Driven Phrase Structure Grammars with Lexical Rules. *Computational Linguistics 17(3)*. 301–314.

Chen, Ch.-M, and Chung, Ch.-J. 2008. Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle. Computers & Education 51 (2008). 624 – 645.

Chen, Ch.-M. and Dug L.-J. 2008. Personalized web-based tutoring system based on fuzzy item response theory. Expert Systems with applications. 34 (2008). 2298 – 2315.

Chun, D. M. & Plass, J. L. 1996. Facilitating Reading Comprehension with Multimedia. *System.* 24:503–19.

Cobb, T. 1999. Breadth and Depth of Lexical Acquisition with Hands-On Concordancing. *Computer Assisted Language Learning* 12(4):345–60.

Dearman, D. and Troung, K. 2012. Evaluating the implicint acquisition of second language vocabulary using a live wallpaper. CHI' 12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1391 – 1400.

DeKeyser, R. M., and Criado-Sánchez, R. 2012. Practice in second language instruction. In C. A. Chapelle (Ed.), The Encyclopedia of Applied Linguistics. Oxford: Wiley-Blackwell.

Diedrichsen, E. 2014. A role and reference grammar parser for German. In Brian Nolan and Carlos Periñán-Pascual (Eds.). *Language Processing and Grammars. The role of functionally oriented computational models*, pp. 105 - 142. Amsterdam: John Benjamins.

Dyk, J. In Press. How do Hebrew Verbs Differ? A Flow Chart of the Differences. *Contemporary Examinations of Classical Languages (Hebrew, Aramaic, and Syriac): Valency, Lexicography, Grammar, and Manuscripts.* Perspectives on Linguistics and

Ancient Languages 5, ed. T. M. Lewis, A. Salvesen, and N. Al-Jeloo. Piscataway, NJ.: Gorgias.

Edge, D., Fitchett, S., Whitney and Landay, J. 2012. MemReflex: Adaptive Flashcards for Mobile Microlearning. Retrieved http://research.microsoft.com/en-us/people/daedge/memreflex.pdf on 2014-12-21.

Ellis, N. C. 1994. Vocabulary Acquisition: The Implicit ins and outs of Explicit Cognitive Mediation. In *Implicit and Explicit Learning of Languages*, Ellis, N. C., ed. London: Academic Press, pp. 211–82.

Ellis, R. (1998b). Teaching and research: Options in grammar teaching. *TESOL Quarterly, 32,* 39-60.

Faber, P, and Mairal, R. 1999. Constructing a Lexicon of English Verb. Berlin: Mouton.

Fellbaum, C., ed. 1998 *Wordnet. An electronic lexical database*. Cambridge, MA.: MIT Press.

Fogg, B.J. 2003. *Persuasive Technology: Using Computers to Change What We Think and Do*. San Francisco, CA: Morgan Kaufmann.

Gottschalk, Judith. 2010. Storage of linguistic knowledge in the mental lexicon: An approach within Role and Reference Grammar. *ITB Journal, Issue 19,* 20 – 45. http://www.itb.ie/files/journal/issue-19.pdf

Gottschalk, Judith. 2012a. On the computability of Role and Referenc Grammar: A new approach to the analysis of three-place predicates and the hierarchical lexicon. *ITB-Journal Issue 22*, 124 – 192. <http://www.itb.ie/files/journal/issue-22.pdf>

Gottschalk, Judith. 2012b. The Persuasive Tutor: a BDI Teaching Agent with Role and Reference Grammar Language Interface. *ITB Journal 23*, 31-51. <http://www.itb.ie/files/journal/issue-23.pdf>

Gottschalk, Judith. 2014. Three-place predicates: A computational approach. In Language Processing and Grammars. The role of functionally oriented computational models, Brian Nolan and Carlos Periñán-Pascual (Eds.) 79 - 104. Amsterdam: John Benjamins.

Gottschalk, Judith. Accepted. On the application of conceptual graphs in RRG – First steps towards a functional computational processing model.

Gottschalk, Judith. ms. Learning Journey Online: Persuasive Feedback for Learners.

Gottschalk, Judith and Winther-Nielsen, Nicolai. 2013. Persuasive skill development: On computational surveillance strategies for modeling learning statistics with PLOTLearner. In *IWEPLET 2013 Proceedings*. Reinhold Behringer and Georgina Sinclair (Eds.) 109-116. Retrieved 14 January 2014 from

http://www.lulu.com/shop/reinhold-behringer-and- georgina-sinclair/iweplet-2013-proceedings/paperback/product-21189131.html.

Groot, P. J. M. 2000. Computer Assisted Second Language Vocabulary Acquisition. *Language Learning and Technology* 4(1):60–81.

Gu, Y. & Johnson, R. K. 1996. Vocabulary Learning Strategies and Language Learning Outcomes. *Language Learning* 46(4):643–79.

Guest, Elizabeth. 2008. Parsing for Role and Reference Grammar. In *Investigations of the Syntax -Semantics-Pragmatics Interface*, Van Valin, Robert D., Jr. (ed). 435 – 453. Amsterdam: John Benjamins.

Hooper, C., Preston, A., Balaam, M., Seedhouse, P., Jackson, D., Pham, C., Ladha, Cassim, Kadha, K., Plötz, T. and Olivier, P. 2012. The french kitchen: task-based learning in an instrumented kitchen. UbiComp '12 Proceedings of the 2012 ACM Conference on Ubiquitous Computing. ACM: New York, pp.193-202.

Hamari, J., Koivisto, J., and Sarsa, H. 2014. Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *proceedings of the 47thHawaii International Conference on System Sciences*, Hawaii, USA, January 6-9. 2014

Jackendoff, R. 2003. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. New York: Oxford University Press.

Kang, S-H. 1995. The Effects of a Context-Embedded Approach to Second-Language Vocabulary Learning. *System* 23(1):43–55.

Mairal-Usón, R. 2003. Why the notion of lexical template? *Anglogermánica online* 2002–1. Retrieved 21 December 2014 from http://www.uv.es/anglogermanica/ag02/Mairal.htm

Laurillard. D. (2012). T*eaching as a Design Science: Building Pedagogical Patterns for Learning and Technology*. London: Routledge.

Leffa, V. J. 1992. Making Foreign Language Texts Comprehensible for Beginners: An Experiment with an Electronic Glossary. *System* 20:63–73.

Lyman-Hager, M. A. 2000. "Bridging the Language-Literature Gap: Introducing Literature Electronically to the Undergraduate Language Student." *CALICO Journal* 17(3):431–52.

Metsämuuronen, J. 2013. Effect of repeated testing to the development of Biblical Hebrew language proficiency. *Journal of Educational and Psychological Development,* 3(1). Retrieved October 10, 2013 from http://www.ccsenet.org/journal/index.php/jedp/article/view/19582.

Murtagh, Irene. Towards a Linguistically Motivated Irish Sign Language Conversational Avatar. *ITB Journal*, Issue 19. 72 – 101.

Nagy, W. 1997. On the Role of Context in First- and Second-Language Vocabulary Learning." In *Vocabulary: Description, Acquisition and Pedagogy*, Schmitt, N. & McCarthy, M., eds. Cambridge: Cambridge University Press, pp. 64–83.

Nation. I. S. P. 2001 *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nolan, Brian. 2004. *RRG 2004 Book of Proceedings*, Brian Nolan (Ed.) 194-223. Retrieved 21 December 2014 from <http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg.html>

Nolan, Brian. 2004. First steps towards a computational RRG. In *RRG 2004 Book of Proceedings*, Brian Nolan (Ed.) 194-223. 14 January 2014 http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg.html (14 January 2014).

Nolan, Brian. 2013. Extending a lexicalist functional grammar through speech acts, constructions and conversational software agents. Plenary talk presented at the 2st International Conference on Meaning and Knowledge Representation in LCM and FunGramKB. Facultad de Ciencias Económicas y Empresariales de Granada. Universidad de Granada, in Granada, Spain.

Nolan, Brian and Periñán-Pascual, Carlos. 2014. *Language Processing and Grammars. The role of functionally oriented computational models.* Amsterdam: John Benjamins.

Nolan, Brian and Yasser Salem. UNIARAB: A Universal Machine Translator System for Arabic Based on Role and Reference Grammar. 14 January 2014 <http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg.html> (14 January 2014).

Ogata, H. and Yano, Y. Context-aware support for computer-supported ubiquitous learning. In *Proc. WMTE 2004*, IEEE Computer Society (2004), 27-34.

Periñán-Pascual, C. 2013 Towards A Model of Constructional Meaning for Natural Language Understanding. Pp. 205-230 in *The Role of Constructions in a Functional Grammar*, ed. B. Nolan and E. Diedrichsen. Amsterdam/Philadelphia: John Benjamins.

Periñán-Pascual, C. and Mairal-Usón, R. 2009 Bringing Role and Reference Grammar to Natural Language Understanding. *Procesamiento del Lenguaje Natural* 43: 265-27

Periñán-Pascual, Carlos and Francisco Arcas-Túnez. 2010. The architecture of FunGramKB. In Proceedings of the 7th International Conference on Language Resources and

Evaluation, 2 667--2674. Malta: European Language Resources Association.

Periñán-Pascual, Carlos and Fransisco Arcas-Túnez. 2014. The implementation of the CLS Constructor in ARTEMIS. A computational approach. In *Language Processing and Grammars. The role of functionally oriented computational models*, Brian Nolan and Carlos Periñán-Pascual (Eds.) 165 - 196. Amsterdam: John Benjamins.

Petersen, Ulrik. 2003. Automatic Lexicon-based Ontology-creation A methodological study. Manuscript. Retrieved 14 January 2014 from http://people.hum.aau.dk/~ulrikp/studies.html (14 January 2014).

Petersen, U. 2003 Automatic Lexicon-based Ontology-creation. A methodological study. Unpublished 9th semester project report. Available at http://www.hum.aau.dk/~ulrikp/studies.html.

Petersen, U. 2004 *Creation in graphs: Extracting conceptual structures from Old Testament texts*. MA thesis, University of Aalborg, Department of Communication. Published in: *Impact* - an Electronic Journal on Formalisation in Media, Text and Language - Impact Theses. Available at http://www.impact.aau.dk/theses.html and http://ulrikp.org/MA/

Petersen, Ulrik. 2007. Genesis 1:1-3 in Graphs: Extracting Conceptual Structures from Biblical Hebrew, SEE-J Hiphil Volume 4, 14 January 2014 http://www.see-j.net/hiphil (14 January 2014).

Pinker, S. 1994. *The Language Instinct: How the Mind Creates Language*. NY: HarperCollins.

Ponniah, J., R. 2011. Incidental Acquisition of Vocabulary by Reading . Retrieved 21 December 2014 fromhttp://www.readingmatrix.com/articles/april_2011/ponniah.pdf at 2014-12-21.

Quast, Jennifer Elizabeth. 2010. Using processing instruction to teach Biblical Hebrew grammar. Hebrew Union College - Jewish Institute of Religion (Ohio).

Rogers, Yvonne, Helen Sharp and Jenny Preece. 2011. *Interaction Design: Beyond Human Computer Interaction*. Third Edition. New York: Wiley and Sons.

Salem, Yasser. 2009. A generic framework for Arabic to English machine translation of Simplex sentences using the Role and Reference Grammar linguistic model. 14 January 2014 <http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg.html > (14 January 2014).

Sanders, Elizabeth, B.-N. 2005. Information, inspiration and co-creation, in: Sixth

International Conference of the European Academy of Design, March 29–31, 2005. University of the Arts, Bremen, Germany.

Schmitt, N. 1997. Vocabulary Learning Strategies. In *Vocabulary: Description, Acquisition and Pedagogy*, Schmitt, N. & McCarthy, M., eds. Cambridge: Cambridge University Press, pp. 199– 227.

Skovenborg, D. ms. A new PLOT for Biblical Hebrew Learning.

Sowa, John F. 2000. Knowledge Representation: Logical, Philosophical, and Computational Foundations. Pacific Grove, CA: Brooks/Cole Thomson Learning.

Thompson, J. ms. The Bible Sense Lexicon: WordNet Theory Adapted for Biblical Languages.

Van Valin, Robert D, Jr. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.

Van Valin, Robert D, Jr. 2007. The Role and Reference Grammar Analysis of Three-Place Predicates. *Suvremena Lingvistika 33.1(63)*.31 – 64.

Van Valin, Robert D., Jr. 2008. Investigations of the syntax-semantic-pragmatics interface. Amsterdam: Benjamins.

Van Valin, Robert D., Jr., und Randy J. LaPolla. 1997. *Syntax: Structure, meaning and function*.

Cambridge: Cambridge University Press.

Webb. S. 2008. The effects of context on incidental vocabulary learning. Accessed at http://nflrc.hawaii.edu/rfl/October2008/webb/webb.html on 2014-12-21.

Wilson, Chris. 2009. Lex: a Software Project for Linguists. Technical Report SEE-J Hiphil 6 14 January 2014 <http://www.see-j.net/index.php/hiphil> (14 January 2014).

Available at http://www.see-j.net/index.php/hiphil/article/view/60

Winther-Nielsen, Nicolai. .2009. Biblical Hebrew parsing on display: The Role-LexicalModule (RLM) as a tool for Role and Reference Grammar. SEE-J Hiphil 6 14 January 2014 [http://www.see- j.net/index.php/hiphil], 1- 51.

Winther-Nielsen, N. 2013a. PLOTLearner as Persuasive Technology: Tool, Simulation and Virtual World for Language Learning. In R. Behringer & G. Sinclair (Eds.), pp. 21-28.

Winther-Nielsen, N. 2013b. PLOTLearner for a Corpus of the Hebrew Bible: The Case for Repurposing in Language Learning. In R. Behringer & G. Sinclair (Eds.), pp. 53-60.

Winther-Nielsen, N. ms. The Corpus as Tutor: Data-driven Persuasive Language Learning. Accepted for publication in *Digital Humanities Quarterly*.

Winther-Nielsen, N. ms. Global tutoring by the Hebrew Bible: - The PLOTLearner Achievement. Accepted for publication in HIPHIL Novum vol 8 (2013), issue 2.

Winther-Nielsen, N. & Hery, A. 2013. The Survey of the Graduate Students. In Nicolai Winther-Nielsen, ed, *Evaluation of PLOTLearner 2. Deliverable for EuroPLOT,* 41-59.

Winther-Nielsen. Ms. Corpus-driven Valence: *Give* and the meaning of נתן *nātan* in Genesis.

Appendix

From: **Lisann Kuenkel** lisann.kuenkel@hpsl.uni-freiburg.de
Subject: RRG 2013
Date: 4 Nov 2014 09:45
To: Judith Gottschalk gottschalk.judith@gmail.com

LK

Dear Judith,

I am happy to let you know you that your paper has been accepted for publication in the RRG 2013 proceedings.
You'll find the review and edited version of your text as well as the stylesheet attached to this mail. Please revise your paper and send it back to me (lisann.kuenkel@hpsl.uni-freiburg.de) before Monday, 31 November, 2014.
May I kindly ask you to send me all the figures as separate files in addition to a copy of your article in word as well as in pdf.

Best wishes,
Lisann


edit anonymous.docx


review anonymous.docx


Stylesheet.docx