DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

„Measuring syntactic complexity in the academic writing of English students at the University of Vienna“

verfasst von / submitted by

Thomas Walter

angestrebter akademischer Grad / in partial fulfillment of the requirements for the degree of

Magister der Philosophie (Mag.phil.)

Wien, 2017 / Vienna, 2017

Studienkennzahl lt. Studienblatt/ degree programme code as it appears on the student record sheet:	E 190 423 344 A
Studienrichtung lt. Studienblatt/ degree programme as it appears on the student record sheet:	Lehramtsstudium UF Chemie UF Englisch
Betreut von / Supervisor:	Univ. Prof. Mag. Dr. Christiane Dalton-Puffer

Acknowledgements

First of all, I want to thank Professor Christiane Dalton-Puffer for her great guidance and the manifold ways of support throughout this project. Above all, it was her support during my seminar paper on a similar topic and her fabulously conducted seminar on learner language that greatly influenced my decision to write about this topic and at the English Department.

Special thanks goes to all participants and Professors at the department who sacrificed some of their precious (seminar) time for the data collection and their interest in my topic. I also want to thank Professor Armin Berger, Bernhard, David, Stefanie and Sophie for their suggestions and their (last minute) feedback on some of the statistical or textual parts of the thesis.

On a personal level, I want to thank my parents Rosi and Gerhard for supporting me whenever possible. I would not have tried nor accomplished a lot of things without knowing that there is always a place for me where I am welcomed and loved. Most importantly, they equipped me with an influential and valuable thing in my life that makes things like this thesis possible: a carefree and colorful childhood full of great and small adventures near and far from home.

I also want to thank my fantastic three ;-) sisters Karin, Victoria and Magdalena for their constant encouragement during the last months. Not to mention that without their tireless, straight and cheeky nature my life would be only quarter as much fun and adventurous!

I am also most grateful to my grandparents Anna and Josef, who always valued and encouraged further education and the passing on of knowledge in our family. Thanks goes also to my aunts, uncles and cousins from both family sides. All of my relatives contributed in their own special way to raise and nourish my interest in science, language and, particularly, in teaching.

A very special thanks goes to the people who supported me through their contributions to a supportive, enriching and reflective living environment, their advice in all kinds of ways or by just being there and spending time with me outside the library throughout the last months. Thank you Ria, Stefanie, Oyunsuren, Kathi, Patrick, Daniel, Börni, Georg, Karl, Matthias, Maxi, Janine, Andi, Ylvie, Maria & Bernhard!

I also want to thank everyone from Team Henriette for the great working environment, switching shifts in the last second, the weekend office space, the cat calendar and the great understanding for my less energetic-than-usual behavior during my last months there!

In representation for all people I was and still am allowed to learn from and the people I forgot to mention here, I want to thank my former teacher Edgar Neukirchner. He raised my understanding of the social, economic and technical implications of the open source concept. I strongly believe that the idea of open source is, besides language, one of the most powerful ways to fruitfully change the way we live and think together.

Table of contents

Index of figures....ii
Index of tables....iii
List of abbreviations....iv

1 Introduction....1

2 The writing process in a second language....4

2.1 (Cognitive) resources for writing....4
2.2 A cognitive model for speaking....9
2.3 Cognitive demands on the L2 writer....13

3 Complexity in the L2 context....17

3.1 Defining Complexity, Accuracy and Fluency....18
3.2 Trade-off hypothesis....22
3.3 Cognition hypothesis....23
3.4 Dynamic System Theory....25

4 Defining Complexity....28

4.1 L2 complexity....28
4.2 Syntactic complexity....31
4.3 Operationalizing syntactic complexity....33
4.4 Syntactic complexity in L2 academic writing....39

5 Research questions....43

6 Materials and methods....46

6.1 Data collection and modification....46
6.2 Participants and sample information....47
6.3 Syntactic complexity measures....50
6.4 Analysis software....51

7 Results....55

7.1 Q1: Change between the two seminar papers....55
7.2 Q2: Influence of writing courses....59
7.3 Q3: Influence of previous secondary education....62

8 Discussion....65

8.1 Q1: Change between two seminar papers....65
8.2 Q2: Influence of writing courses....67
8.3 Q3: Influence of previous secondary education....68

9 Conclusion & implications for further research……70

References..72

Appendix...81

1 Info-sheet and questionnaire...81
2 Learner background information...84
3 Script for editing the input files..86
4 Example paragraphs before and after modifications...87
5 Texts used for L2SCA comparison...88
6 Parse tree example used for comparison...100
7 Manual and L2SCA annotations...100
8 Summary of manual and L2SCA comparison..113
9 Measurement results...114
10 Confirmation of paired sample t-test assumptions..119
11 Confirmation of independent sample t-test assumptions.......................................120
12 Confirmation of ANOVA assumptions..121
13 Paired sample t-test for L1 German samples...122
14 One-way ANOVA result tables...122
15 Abstract (English)..123
16 Abstract (German)..124

List of figures

Figure 1: Levelt's blueprint for the speaker...10
Figure 2: A taxonomy of complexity constructs...28
Figure 3: Construct definition of grammatical complexity...31
Figure 4: A multidimensional representation of syntactic complexity.......................33
Figure 5: Boxplots for all seven measures between PS paper and SE paper.............56
Figure 6: Boxplot for Δ-MLC and Δ -CP/C across number of writing courses.........61
Figure 7: Boxplots for all seven PS paper measurement results between AHS and BHS........63

Index of tables

Table 1: Descriptive statistics of the corpus....47
Table 2: Descriptive statistics of the learners....47
Table 3: Definition of measures according to sub-constructs....50
Table 4: Sequencing unit identification between annotator and L2SCA....52
Table 5: Pearson's correlation coefficient (r) between annotator and L2SCA measures....54
Table 6: Difference of syntactic complexity between PS paper and SE paper for all samples. 58
Table 7: Group statistics of the one-way ANOVAs for Δ-MLS and Δ-CP/C....60
Table 8: Differences of syntactic complexity between AHS PS papers and BHS PS papers....64
Table 9: Detailed sample description....84
Table 10: Comparison of annotations for PS paper of sample 1....101
Table 11: Comparison of annotations for SE paper of sample 1....103
Table 12: Comparison of annotations for PS paper of sample 28....105
Table 13: Comparison of annotations for SE paper of sample 28....107
Table 14: Comparison of annotations for PS paper of sample 57....109
Table 15: Comparison of annotations for SE paper of sample 57....111
Table 16: Summary of annotations for all 6 samples....113
Table 17: Correlation data between manual and L2SCA measures....113
Table 18: Results for PS samples....114
Table 19: Results for SE samples....116
Table 20: Differences between SE and PS measurement....118
Table 21: Kolmogorov-Smirnov results for paired sample test values....119
Table 22: Kolmogorov-Smirnov test results for AHS group values....120
Table 23: Kolmogorov-Smirnov test results for BHS group values....120
Table 24: F-test results for equality of variances between AHS and BHS samples....120
Table 25: Kolmogorov-Smirnov test results for ANOVA residuals....121
Table 26: Levene test results for homoscedasticity of ANOVA group variances....121
Table 27: Difference of syntactic complexity for samples with only German as L1....122
Table 28: One-way ANOVA results for ΔMLC with factor no. of courses....122
Table 29: One-way ANOVA results for ΔCP/C with factor no. of courses....122

List of abbreviations

BCS	Bosnian, Croatian or Serbian
CAF	Complexity, accuracy and fluency
CFT	Common Final Test
CLIL	Content and Language Integrated Learning
CP/C	Coordinate chrases per clause
DC/TU	Dependent clause per T-unit
DST	Dynamic System Theory
EAP	English for academic purposes
EFL	English as a foreign language
L1	First language
L2	Second language
L2SCA	Second language syntactic complexity analyzer
SD	Standard deviation
SLA	Second language acquisition
MLC	Mean length of clause
MLS	Mean length of sentence
MLTU	Mean length of T-unit
NFE/C	Non-finite elements per clause
PS	Proseminar
SE	Seminar
TBLT	Task based language teaching
TU/S	T-unit per sentence
95%-CI	95 per cent confidence interval

1 Introduction

Second language (L2) complexity is seen as one important construct to investigate learner language. Together with accuracy and fluency, complexity is argued to reveal important information about the state of the L2 system in the learner's mind. As such, complexity, accuracy and fluency (CAF) figure as variables to assess overall L2 proficiency, to trace L2 development or serve as a descriptor for L2 task outcomes in Task Based Language Teaching (TBLT) research (Ortega 2012: 128). Syntactic complexity represents one specific sub-dimension within the multidimensional construct of L2 complexity. Attempts to define syntactic complexity in a reasonable way within the CAF framework and as a distinct language property proved to be challenging (Bulté & Housen 2012: 21–22; Ortega 2015: 90). Notwithstanding these complications, recent advances in the field of Second Language Acquisition (SLA) helped to consider various aspects that might play a role in the syntactic development of L2 writers. While cognitive approaches to SLA provide valuable models of syntactic development in the L2 writer, TBLT research and corpus linguistics aim to elaborate measurement procedures and construct definitions to assess syntactic complexity in L2 production. For this reason, it is important to know which factors might influence syntactic complexity in the L2 learner. Examples for variables that influence syntactic complexity are the learner's L1, time and amount of L2 writing instruction, the institutional setting, the proficiency level or the writing topic (Lu 2011; Lu & Ai 2015; Mazgutova & Kormos 2015; Yang, Lu, & Weigle Cushing 2015). Additionally, text genres have been shown to alternate syntactic complexity outcomes. Despite not being investigated so far, it is also assumed that previous genre exposure has an influence on syntactic complexity measures (Lu 2011: 48; Ortega 2015: 88).

This thesis is based on these findings and intends to provide an example for a way to measure syntactic complexity in academic papers, in this case of students at the English department of the University of Vienna. It explores the application of one specific construct to measure syntactic complexity in writing and attempts to provide, in this way, an overview of the syntactic development of L2 learners in a very specific academic setting. The obtained findings should yield useful information about learner's academic writing at the English department of the University of Vienna. This information can also be used as a starting point for further research at the department or across institutions. Effects such as writing instruction or proficiency level have already been investigated to some extent with respect to syntactic complexity in academic writing. Most studies, however, have investigated syntactic complexity in learners who learned English as an L2 in an English speaking country, but there are few studies about syntactic development of English students in non-English speaking countries. Furthermore, the influence of previous secondary education and genre exposure have not yet been considered in the syntactic development of L2 learners. Based on the previously outlined influences on syntactic development, three research questions will be addressed:

(Q1) *Does syntactic complexity change in the academic writing of students at the English department of the University of Vienna between their first linguistic seminar paper and their last linguistic seminar paper?*

(Q2) Does the number of academic writing courses taken between two linguistics courses account for differences in syntactic complexity?

(Q3) Do dimensions of syntactic complexity differ in the academic writing of EAP learners depending on the type of secondary school they attended in Austria?

These research questions will be discussed in chapter 5 based on the theoretical part, which will be put forward in the following chapters. The theoretical part will help to relate the measures that were used to assess syntactic complexity to cognitive aspects of language because some cognitive components of the learner’s L2 system have been argued to correspond with syntactic complexity. Therefore, chapter 2 will start by explaining the writing process on behalf of the resources that are necessary for writing and connects them with the cognitive demands put on the L2 writer. In chapter 3, L2 complexity will be discussed as an integral part of the multidimensional CAF construct in the light of recent theories of CAF. Chapter 4 will outline the various sub-constructs that comprise L2 complexity with a focus on the operationalization of syntactic complexity measures to tap aspects of L2 complexity. Moreover, chapter 4 concludes the theoretical part with a review of recently conducted research related to syntactic complexity in academic writing. The methodological part will attempt to answer the three research questions which are broadly discussed in chapter 5. In the materials and methods section that constitute chapter 6 the elicitation of the learner data and the tools that have been used for the analysis are described. This chapter also includes a sample description with detailed learner background information and a validation of the analysis software used. Subsequently, in chapter 7 the results of the statistical data analysis to answer the research questions are presented. While the results are discussed in chapter 8 in the light of previous research outcomes, chapter 9 draws some conclusions about the study outcome and points out implications for further research.

2 The writing process in a second language

2.1 (Cognitive) resources for writing

Writers can draw on resources in different ways than speakers do in order to accomplish their goals, because in most cases there is more time available to generate ideas or text structures and the span for conscious revision is longer (Williams 2012: 322). On the one hand, writing relies on cognitive resources that are used both during planning and the actual writing process. On the other hand, writers also rely on non-cognitive resources provided by the writer’s environment (Hayes & Olinghouse 2015: 482–7). Both will be briefly discussed in the following sub-sections.

Task environment

Writers use task materials such as prompts for writing tasks as well as other input from their immediate physical environment as non-cognitive resources for their writing. Dictionaries, peers or a teacher providing feedback and the immediate surroundings such as the architecture of the room in which a text is produced would all be examples for non-cognitive resources. All these are part of what is often termed *task environment*, which is directly involved in the actual cognitive process of writing. Two other important parts of the task environment are the transcribing technology and the text already written by the writer. Firstly, transcription technology such as pen and paper or a computer keyboard partially influences the cognitive demands during writing, depending on the degree of practice with the technology used. At some point during language learning, however, transcription will be automatized to such a degree that it uses minimal cognitive resources (McCutchen 1996: 319). Secondly, the text written so far has been shown to not only influence revision, but also support the generation of new text segments. It can be understood as a short-term memory extension (Chenoweth & Hayes 2003: 115).

Reading

As already mentioned, the already written text and the task-environment are part of the non-cognitive resources, but the writer has to make use of cognitive resources to integrate information from the task environment into the writing process. Crucial cognitive resources for language production are working memory, attention, long-term memory and reading. Hence, to utilize the already written text the writer relies on *reading*. Two types of reading can be distinguished, depending on the goals of the writer. One type is *reading for editing and revising*, which directs attention towards grammar, spelling or the general structure of the recently written text. Moreover, it includes repeated reading of parts of the text and can also trigger the production of new text, ideas and syntactic structures in some way (Alamargot, Dansac, Chesnet, & Fayol 2007: 29). The second type of reading is when writers are composing with the help of other text sources. This is described as *reading for gist* to extract required information for the text to be written. The extracted information can then be collected on a written plan that is part of the task environment and serves as help to structure the text. Otherwise, the information can be temporarily stored in the writer's memory for generating new ideas or integrating them into the text (Hayes & Olinghouse 2015: 371).

Long-term memory

To create new ideas for writing, writers take advantage of another cognitive resource, *long-term memory*. It stores the writer's general world knowledge such as facts and topics, but it is also the place where the writer's language knowledge such as vocabulary, spelling, grammar or orthographic skills is stored (Hayes & Olinghouse 2015: 486). One specialized part of long-term memory is the *mental lexicon* containing information about words in the language(s) a person knows (Levelt 1989: 6). According to Levelt, the mental lexicon retains this information about lexical items in two ways. On the one hand as *lemmas*, representing the information about the meaning and the syntactic environment of a word (11), on the other hand as the *lexical form* of a word, carrying morphological and phonological information about a word (12). Taking the L2 mental lexicon into account, it has been argued that the lexical information for the L2 is usually smaller in its size, elaboration and organization than the L1 mental lexicon (Skehan 2009: 529).

Another form of language information stored in the long-term memory are schemas. *Schemas*, in general, are ways to organize and categorize knowledge and automatized processes for certain aspects of the world. If a schema is automatized, it needs fewer cognitive resources to be processed (Gathercole & Baddeley 1993: 6). An example for a non-language related schema would be opening a door including all cognitive processes involved such as the actual recognition of a door in one's surrounding, identifying the handle and pushing it down. Writing schemas, therefore, include the writer's knowledge about texts genres, cultural influences, their structure and what kind of language has to be used for a certain kind of text in terms of formality or syntactic structure. Writing schemas also represent a rough working plan of what has to be done to produce a certain text type and influence how the different writing processes interact. A very specific example for such a writing schema is writing an academic paper which is argued to be mainly learned after secondary education. However, previous writing experience and access to certain genres, as provided by secondary education, is argued to affect the development of later acquired writing schemas, such as academic writing (Biber, Gray, & Staples 2016: 645). Considering this, writing schemas are important factors influencing the quality of produced texts. Moreover, they can be learned and extended through instruction over a whole lifespan (Hayes & Olinghouse 2015: 483; Ortega 2015: 88).

Overall, information that is stored in long-term memory can be divided into declarative knowledge and procedural knowledge. *Declarative knowledge* comprises information from *semantic memory* and *episodic memory*. Information such as words in the mental lexicon, conceptual knowledge or facts are stored semantic memory and different from experienced events that are stored in episodic memory. It has been argued that information from both memories is expressed differently through language. While the meaning of physical objects or mental concepts in semantic memory is expressed through words, events and event chains from episodic memory are expressed as syntactic elements in form of clauses and clause chains in language (Givón 2009: 11). Information from both kind of memories is described as knowledge 'that' something is. This stands in contrast to 'how' something is done, which is termed *procedural knowledge*. Procedural knowledge entails all knowledge needed to execute cognitive or psychomotoric skills such as opening a door, retrieving words from the mental lexicon or the steps to be taken for writing an essay (De Keyser 2009: 121).

According to cognitive approaches to SLA, this knowledge perspective plays a major role in learning a L2. Broadly speaking, L2 knowledge, once stored as declarative knowledge, can be proceduralized and through practice will be automatized at a certain stage of L2 development; thus costing less mental effort and attention in processing. A grammatical rule, for instance, can be stored in the mind of the learner as declarative knowledge. Eventually, it will become automatized, after frequent use of the rule and its accompanied proceduralization with frequent activation of the respective brain areas (De Keyser 2009: 131). Nevertheless, there are a lot of other factors involved that influence the acquisition of new L2 knowledge, but this particular knowledge distinction will help to explain the following cognitive resources and their interactions.

Attention

The cognitive resource of attention is described as the “ability to maintain focus on a task in the face of distraction” (Hayes & Olinghouse 2015: 486). To elaborate this definition, it is a general assumption of cognitive psychology that a person can only consciously attend to a certain number of sensory inputs or outputs at the same time, i.e. attention is a limited resource. All cognitive tasks require a certain processing time and cognitive effort, unless the performed cognitive tasks are *automated*, meaning they are independent from attention and costing no cognitive effort. Processing time reflects the time spent attending to a cognitive task, while cognitive effort is the amount of free attention momentarily spent on a task (Kellogg 1994: 32). It has been argued that cognitive processes can work in parallel as long as they do not exceed the total amount of available attention (Kellogg 1996: 59). Because writing – as well as speaking – consists of various cognitive processes that often have to be handled simultaneously, it is important for writers to focus their attention on specific processes to produce compositions (Hayes & Olinghouse 2015: 486). At the same time, the role of attention in L2 production processes poses particular complications on the learner’s performance such as more time to retrieve words from memory or more frequent use of repair strategies. This is why writing in the L2 often needs more conscious and selected allocation of attention (Kormos 2014: 196).

Working memory

The last cognitive resource to be discussed is *working memory.* It characterizes a short-term memory system that can be seen as a constantly available work space to the writer and is individually limited in its capacity (Gathercole & Baddeley 1993: 5). The most important part of working-memory is the *central executive*. It is the major component managing information transmissions between other parts of the cognitive system and directing information to more specialized parts of working memory. In the central executive information from other cognitive resources is held, sorted and re-structured to form new ideas. Moreover, the central executive does not only provide required information for all cognitive processes such as decision making, planning or non-language specific cognitive tasks, but it is also involved in retrieving information from long-term memory (McCutchen 1996: 300). Despite the importance of the central executive for language production processes, it has been shown that other parts of working memory in adults are further specialized for speaking and writing.

One such specialized part of the working memory has been termed the *phonological loop* (Gathercole & Baddeley 1993: 8). The phonological loop can be further distinguished into a *phonological short-term store* and an *articulatory rehearsal process*. The phonological short-term store keeps verbal material ready for production and/or revision. Furthermore, the outputs of linguistic encoding processes are temporarily stored and bits of already produced language are available to the writer's attention; thus it plays also a role for reading or naming objects. More importantly, it has been argued that this part is essentially involved in the production of higher order language segments such as phrases, clauses or sentences, which are argued to be distinct dimensions of syntactic complexity. The other part of the phonological loop, the articulatory rehearsal process, is experienced as speaking to oneself and can be observed during writing (Kellogg 1996: 63). Considering this, the articulatory rehearsal process is assumed to be involved in the final planning and evaluation stages before writing a sentence down (60). It should be noted that verbal distraction can influence the writing process, thus indicating mutual use of the same short-term memory resources for both writing and speaking (Chenoweth & Hayes 2003: 116).

Another specialized part of the working memory has been termed visuospatial sketchpad and it is responsible for processing and maintaining input with spatial or visual character (Gathercole & Baddeley 1993: 4). This part of working memory is argued to be mainly used during the planning phase, when the writer often draws on visual imagery or the immediate physical surroundings in creating ideas (Kellogg 1996: 60). Interestingly, reading, in particular for editing and revising, has been shown to draw again on the resources of the phonological loop in the adult, skilled writer (63). However, the phonological loop and the visuospatial sketchpad appear to be limited in their capacity. This means, if higher cognitive demands are required, the central executive of the working memory is accessed to fulfill these demands, but this leaves fewer resources for other cognitive processes (Kellogg 1996: 63). In this context, it has been argued that writing poses higher processing and storing demands on the working-memory than speaking, hence writing requires more attention (McCutchen 1996: 301).

These four cognitive resources of reading, long-term memory, attention, and working memory are argued to be of major importance for writing. Throughout the writing process cognitive resources interact with each other, often simultaneously to structure and fulfill the processing demands of the writing process. Particularly, L2 production processes are likely to use up all attention for cognitive processing, because they often lack automatization and therefore require more cognitive effort or processing time. Another explanation for this is that the retrieval process of L2 knowledge from long-term memory takes longer – taking more processing time – and therefore blocks the retrieval of other items, needed by another process. Once the total amount of attention has been drained, the writer has to prioritize some processes over the others. For instance, transcription will be slowed down and more resources will be dedicated to reading or other processes for the retrieval of information from long-term memory. Therefore, the writer uses planning to structure the allocation of cognitive resources for certain tasks of the production process, thus overcoming to some degree the eventual limitations of resources (Manchón & Roca De Larios 2007: 551). This, however, will become clearer in the following description of Levelt's speaking model.

2.2 A cognitive model for speaking

Despite the fact that there are well established theoretical models for writing (e.g. Hayes 2012), Levelt's model for speaking emphasizes the actual cognitive transformation of ideas into language and the retrieval and organization of linguistic knowledge to form a message. These parts of language production seem to be of central concern for the emergence of syntactic complexity. As far as the theoretical background is concerned, the choice of this model is based on the assumption that some of the cognitive processes and linguistic resources underlying language production are the same in both writing and speaking (Grabowski 1996: 85; Kellogg 1996: 58; Kormos 2014: 197).

Levelt's (1989) componential model for speaking, which can be seen in figure 1, generally dissects the speaking process into four autonomous, specialized components, which are the *conceptualizer*, the *formulator*, the *articulator* and a *speech comprehension system* (9). All of these components are claimed to work in a highly automatic fashion and therefore permit parallel and incremental processing to sustain the fluency of speech in the L1 (2). These processes executed by the listed components, are, however, claimed to be only partially automated when performing in the L2; hence language processing is cognitively more demanding for the L2 speaker (Skehan 2009: 518). Nevertheless, each of these components relies on different cognitive resources to fulfill their functions. For instance, in working memory the outputs of all components can be temporarily stored, waiting to be processed by subsequent components. In this sense, the model is able to account for eventual asynchronicities due to the gradual and parallel processing behavior of each component (Levelt 1989: 24).

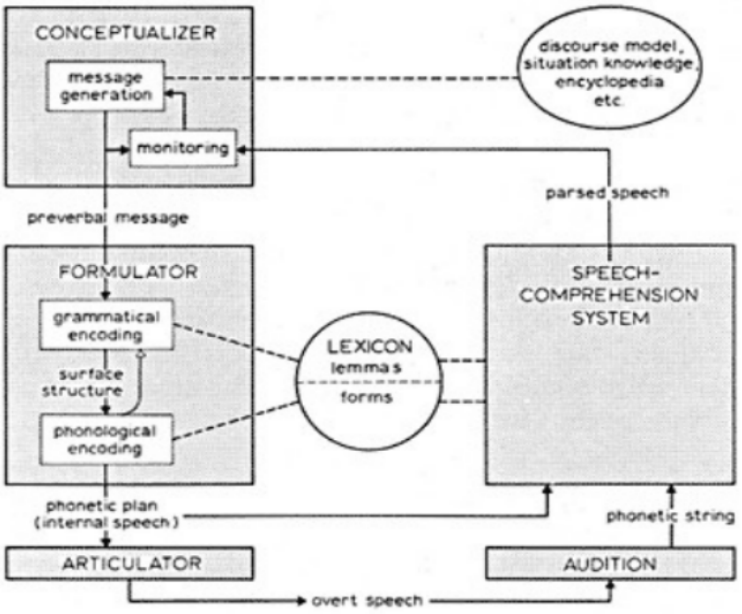

CONCEPTUALIZER
message generation
monitoring
discourse model, situation knowledge, encyclopedia etc.
parsed speech
preverbal message
FORMULATOR
grammatical encoding
surface structure
phonological encoding
LEXICON
lemmas
forms
SPEECH-COMPREHENSION SYSTEM
phonetic plan (internal speech)
phonetic string
ARTICULATOR
AUDITION
overt speech

Figure 1: Levelt's blueprint for the speaker. Boxes represent processing components; circle and ellipse knowledge stores (Levelt 1989: 9)

The first step in generating an utterance is processed by the conceptualizer, as shown in the top left corner of figure 1. This component is triggered by the intention of the speaker to form a message through selecting, ordering and retrieving the relevant knowledge from long-term memory, buffering it in working memory, or analyzing utterances from previous speech acts (9). By doing so, the conceptualizer uses macro-planning and micro-planning to yield the *preverbal message* (10). Levelt refers to the speaker’s act of gathering and sequencing useful information to achieve a communicative goal as macro-planning, while micro-planning is described as the act of shaping information into the right format of the preverbal message to be packed into linguistic knowledge (107). Consequently, the preverbal message contains the results of both planning processes and can be seen as the conceptual information needed to form an utterance. Moreover, with regards to L2 production, it has been argued that the devised preverbal message is conceptually tuned to the requirements of the next stage, the language specific formulator. Therefore, the preverbal message has to be sufficient for the language specific, semantic demands needed by the formulator for the grammatical encoding in the intended language. This means that the preverbal message must contain the information for an utterance in a way that is, for instance, encodable by the spatio-temporal possibilities of a specific language's grammar (106). The basis for this is provided by the argument that every language entails different grammatical systems to encode notions such as space, time or gender. Consequently, this leads to different ways of thinking for language production in different languages and the preverbal message has to be adjusted according to the language specific formulator. Thus, conceptualizing for the L2 formulator has to be learned with greater effort and more attention in contrast to automatization during L1 acquisition (Slobin 1996: 91). Furthermore, with regard to the emergence of complex language structures, it has been argued that more complex concepts in general require higher processing demands at the conceptualizer stage. Although there are differences in the acquisition of such elements between L1 and L2, the need to learn complex language in the L2 is assumed to involve the re-mapping of some concepts already known from the L1 onto corresponding language items in the L2. Thus, "the development of the L2 conceptual system triggers both vocabulary and syntactic development" (Kormos 2011: 54). The preverbal message from the conceptualizer created in this way, however, serves as direct input for the next component, the formulator. Levelt distinguishes between the grammatical and the phonological encoder as two sub-components of the formulator, depicted in figure 1 (1989: 6).

Firstly, the *grammatical encoder* is carrying out the procedures for building syntactic functions according to the information contained in the lemma and for retrieving these lemmas from the mental lexicon. Levelt argues that grammatical encoding is lexically driven (236). Therefore, the processes for grammatical encoding are to a great extent regulated by syntactic information from the lexical items that are stored in the form of lemmas (see 2.1). During the process of grammatical encoding, the lemmas are retrieved from the mental lexicon according to the semantic information from the preverbal message. This gives access to the syntactic information of words stored in the lemma. The specific and specialized syntactic information from the retrieved lemma guides further processes for building what is called a *surface structure,* i.e. a syntactic framework of a message (236). Grammatical encoding of a message can be seen as a main factor involved in the appropriate and fast building of syntactic structures during language production. In this regard, it has also been argued that speaking and writing use the same cognitive resources to syntactically encode the preverbal message into the surface structure (Cleland & Pickering 2006: 195). This surface structure contains abstract information about needed inflections, person or tense for a specific language (Levelt 1989: 164). Apart from language inherent differences in the syntactic encoding process for the L1, it has also been argued that L2 speaking uses a distinct formulator stage (106). During speech production in the L2, the highly automatic behavior of the L1 formulator processes are not given. Procedures for encoding are often not automatized to the same degree in the L2 formulator as in the L1 formulator. Moreover, it is hindered by a smaller and less organized L2 mental lexicon that often does not meet the demands of the preverbal message. Consequently, this leads to a higher degree of repair and replacement processes, raising the processing demands resulting in additional use of cognitive resources to create the surface structure (Skehan 2009: 518).

In the second part of the formulator, the *phonological encoder*, a *phonological plan* is created based on the information contained in the surface structure. For this, the phonological encoder retrieves the lexical forms from the mental lexicon according to the required morphemic and phonological inflections such as plural form or stress (165). The created phonological plan is then buffered in working memory to be available to the articulator for the transformation into *overt speech*. It can, however, also be accessed by the *speech comprehension system* before articulation in form of *internal speech* as this is shown by the arrows in figure 1. For the analysis of internal speech, the speech comprehension system has full access to the mental lexicon where the analyzed message is compared with the lemmas and lexical forms accordingly (13). Similarly, overt speech is picked up by the speech comprehension system from audition and analyzed for its correctness and intended meaning. In both cases, detected errors in production cause the speech comprehension system to put the phonetic message into a feedback loop for re-adjustment or trigger speech repair strategies. This feedback loop starts at the conceptualizer and again goes through all the stages to the articulator, until the intended and correct utterance leaves the mouth of the speaker. At the same time, the obtained information from the audition system serves as new input in form of *parsed speech* provided from the speech comprehension system for the conceptualizer to create the subsequent bits of an utterance (14). The conceptualizer possesses a separate monitoring component to process already articulated speech. This component not only monitors the correct construction of the preverbal message, but also tunes the preverbal message to the already articulated speech, as indicated by the arrows at the monitoring component of the conceptualizer in figure 1. This points towards the cyclical and recurring notion for the immediate production of speech. However, in the long run speaking has been described as a seemingly linear process provoked by the constant time pressure of the modality when compared to writing (Kormos 2014: 197).

Overall, three main stages in speech production can be derived from Levelt’s model: *conceptual preparation* as realized in the conceptualizer, *linguistic formulation* executed by the formulator and the *physical production* resulting in overt speech. The first two processes of conceptualization and formulation are claimed to be similar in writing, while the physical production of articulating has to be substituted with *transcription* to produce output in written form (Schoonen, Snellings, Stevenson, & Van Gelderen 2009: 78). Although it has been argued that phonological information has to be retrieved only partially when writing down isolated words (Bonin, Fayol, & Gombert 1998: 283), other researchers argue that utterances formulated to be written are evident in the speaker’s mind up to the point of internal speech (Chenoweth & Hayes 2003: 100). In any case, language production in the L2 poses a higher cognitive load on the writer or speaker.

2.3 Cognitive demands on the L2 writer

Although the previous sub-chapter indicated similarities between speaking and writing, a writer has more time when writing. Moreover, the permanency of the produced text allows one to draw differently on resources for production. This involves also a different level of importance of certain processes such as planning and revision (Ravid & Tolchinsky 2002: 427). Both are essential, cyclically recurring processes in the course of writing and rely on the same resources as the actual writing process. Consequently, they can be characterized as sub-goals resembling parts of the actual writing process (Hayes 2012: 375).

To begin with, the *planning* process poses different demands on the L2 writer. It should be noted that the term planning in speech production literature refers to the linguistic planning processes which transforms ideas into linguistic units. In contrast, writing literature uses the term planning mainly for the intensive cognitive demands spent on non-linguistic conceptual work preceding the actual transformation of ideas into text (Kellogg 1996: 60). Nevertheless, once a goal for writing has been set and planning has started, writers are likely to make a current plan including sub-goals or arguments that will be included in the written text according to already known *writing schemas* (see 2.1). It has been shown that not only parts of such schemas can be transferred from the L1 to writing in the L2, but also syntactic structures which assemble such schemas are in some cases transferred (Lu & Ai 2015: 25). At the same time, it has been argued that the development of complex structures and expository writing that is expected to happen during the adolescent stage in the L1 are important factors in influencing L2 writing development (Ravid & Tolchinsky 2002: 18). Particularly, previous writing instruction or environmentally determined access to specific genres up to university education in the L1 or in the L2 may have an influence on later acquired complex language structures for academic writing (Berman & Nir-sagiv 2007: 107; Biber et al. 2016: 645). This influence might occur bidirectionally, if writing is practiced more often in the L2 and at higher proficiency levels. Despite the influence of cross-language transfer, definite conclusions about this should be cautiously made, because individual differences of learners seem to affect language transfer in both directions in unpredictable ways (Manchón & Roca De Larios 2007: 578). Moreover, a transfer from the L1 is often hindered because of limited linguistic knowledge in the L2 (Schoonen et al. 2009: 81). Additional research on L2 planning, however, has demonstrated that more proficient L2 writers use more time for planning and consider global issues such as text organization or audience. The revision of syntactic and lexical issues can also become part of this (Campbell 1990: 212). At the same time, L2 writers allocate more time on planning for unfamiliar topics, which allows them to produce more ideas for their writing (Manchón & Roca De Larios 2007: 576). Additionally, the effect of time allowed for planning in L2 writing tasks proved to be beneficial for the use of more complex language forms and seems to promote the restructuring of the L2 system (Williams 2012: 326). This points towards the important role of planning as a tool to structure and consciously allocate cognitive resources. Overall, planning definitely plays an important role for writing in the L2, but aspects of planning can be influenced in many different and individual ways. Therefore, it is difficult to determine to what extent planning is reflected in the writer’s L2 knowledge observed in a composition (Schoonen et al. 2009: 78).

Revision is another important and more specific process, facilitated through the higher amount of time available during the writing process. The revision process in writing can draw on different resources than in speaking and is not only restricted to quick repair strategies as applied in speaking (Schoonen et al. 2009: 88). Revision can also be seen as a compensation strategy for lacking L2 knowledge. It has been shown, however, that revision declines with the level of language proficiency and advanced L2 writers allocate more revision time on improving the quality of an argument or textual coherence (Rosa M. Manchón, Roca De Larios, & Murphy 2009: 112).

It has been shown that writers usually interrupt the transcribing process not only to revise the already written text or to obtain feedback, but also to use it as a support for producing the next part of a sentence. This indicates that planning and revision are closely connected processes (Hayes 2009: 77). Moreover, it illustrates the cyclic, recursive character of the writing process. Despite the peculiarities of revising and planning, writing, similarly to speaking, can be divided up into conceptual preparation, linguistic formulation and physical production (Schoonen et al. 2009: 78). Issues concerning conceptualization in the L2 have been already addressed in the description of Levelt's model and in the planning part of this sub-chapter.

The formulation process, however, seems to consume the highest proportion of time during L2 writing, although processing time declines with improving L2 proficiency (Manchón et al. 2009: 108). The reason for this could be that the sub-processes of formulation in the L2 are often not automatized to the same degree as in the L1. As a result, formulation in the L2 consumes more cognitive resources, because the processes first have to be learned under conscious attention and are only partially automatized (De Keyser 2009: 131). This lack of automatization results in fewer cognitive resources available and therefore parallel processing might suffer in the L2 (Kormos 2014: 196). Consequently, this also leads to a higher cognitive load for the speech comprehension system, ending in more revision, repair and replacement. Less proficient writers, therefore, often avoid complex structures or forms as a compensation strategy (Schoonen et al. 2009: 82). Another way to compensate for the lack of automatization, which does not compromise the complexity of a composition, could be to allocate more time for the retrieval of lexical items from the mental lexicon. This would also explain the high proportion of time spent on formulation in the L2 (Kuiken & Vedder 2011: 92). Nonetheless, it is assumed that with increasing language experience lexical retrieval in the L2 needs less attention. Consequently, attention can be used by the conceptualizer and the formulator to process more complex grammatical structures with higher accuracy. Accordingly, the learner has to make less use of revisions at sentence level and can concentrate on the overall structure of the text (Chenoweth & Hayes 2001: 94). This is also reflected in decreasing time dedicated to formulation and revision at intermediate and higher proficiency levels, while the time allocated for planning increased with experienced L2 writers. To illustrate, it has been shown that in L2 writing the double amount of time is spent on formulation problems during writing compared to L1 writing (Roca de Larios, Manchón, & Murphy 2006: 107). Given these points, linguistic experience seems to facilitate the use of syntactically complex structures during formulation in the L2. It promotes this by increasing the amount of available lexical items to formulate ideas and improving their retrieval rate from the mental lexicon (Chenoweth & Hayes 2001: 94).

Similarly, these problems apply to orthographic encoding in the L2 which is also associated with higher cognitive demands and less effective processing. This is caused by different spelling rules or sometimes even transcription into different graphemes (Schoonen et al. 2009: 80). Although orthographic encoding depends on linguistic knowledge encoded by the formulator stage, its automatization is likely to be already developed to a certain degree by the time one starts to write in the L2 (81). Moreover, it is argued that transcription as opposed to formulation will be automatized relatively quickly to such a degree that it costs minimal and manageable cognitive resources in the L2 (McCutchen 1996: 319). Nevertheless, transcription also relies on practice in the used transcription technology (i.e. writing with pen on paper, typing on a computer keyboard, texting on the phone etc.), which is not always given for every individual learner in the same way as the daily use of articulation.

To conclude, this rough outline demonstrates that the quality of L2 writing is more dependent on the amount of L2 knowledge and the degree of automatization of L2 specific cognitive processes than in the L1. Although planning and revision can compensate for some deficits in L2 knowledge and L2 skills, this chapter illustrated that the conceptualization and formulation processes in the L2 are cognitively more demanding (Schoonen et al. 2009: 94). Despite other factors such as the task environment or planning time, L2 knowledge and the resulting L2 performance is to a certain degree reflected in a composition. The next chapter will discuss one way to assess L2 knowledge of a written product by using the three constructs of complexity, accuracy and fluency.

3 Complexity in the L2 context

The construct of complexity is viewed as one part of a multidimensional framework in SLA research that consists of the three major components: complexity, accuracy and fluency (CAF). It is argued that CAF are able to empirically capture the multicomponential nature of L2 knowledge as a reflection in the learner's L2 production (Housen & Kuiken 2009: 461; Skehan 2014b: 2). As such, they are assumed to reflect different cognitive processing areas involved in L2 production, therefore CAF are seen as distinct variables and areas of L2 performance. Consequently, CAF should always be considered together in theoretical descriptions and in analyses of learner language (Housen, Kuiken, & Vedder 2012: 3).

This insight was not only gained from earlier developmental index studies which attempted to establish an index to objectively measure L2 development on the grounds of CAF measures (Larsen-Freeman 1978: 440, 1983: 287), but also through data from psycholinguistics and cognitive psychology. The latter research fields provide alternative approaches that help researchers to look at SLA from a processing-oriented perspective and identify CAF as the three main components of L2 performance (Skehan 1996: 46). As a result of this cognitively driven approach, hypotheses that aimed to describe an interrelationship between the three elements of CAF arose from task-based SLA research (Robinson 2001; Skehan 1996, 1998). The Cognition Hypothesis and the Trade-off Hypothesis are two rivaling theories which try to explain how CAF might interact during L2 production. Both hypotheses originate from task-based learning research and their validation is an ongoing point of discussion in this field (Robinson 2011a; Skehan 2014b).

Dynamic Systems Theory[1] (DST) is another, recently emerged approach to SLA which uses CAF as indices to investigate individual L2 development and therefore has been partly influenced by the outcomes of L2 developmental index studies. The theoretical base for this approach has been adopted from other natural science fields and views language itself as a dynamic and self-organizing system (Beckner, Bybee, Christiansen, Croft, et al. 2009: 2). In contrast to the cognitive processing perspective, DST takes a descriptive perspective towards L2 development to SLA. In this, DST places itself in the supra-disciplinary space of other theories – including cognitive linguistics – that primarily serves as a tool for description of empirical data to subsequently deduce patterns and models to explain the development of CAF (Cameron & Larsen-Freeman 2007: 15–6).

---

[1] Although this research approach comprises more than one name with slightly different focuses such as Chaos Theory, Complex Adaptive System Theory or Dynamic Systems Theory, they all take a similar perspective onto SLA. Because these differences do not influence the argumentation of this thesis and for the sake of convenience, they will be summarized under the term Dynamic System Theory (DST) (Dörnyei 2009: 99–100).

Overall, each of the three approaches to be discussed views syntactic complexity as one dimension of language complexity and therefore as an integral part of CAF. Before fully turning to theories about CAF, the first part of this chapter starts with discussing recent definitions of the terms complexity, accuracy and fluency. There are several different definitions for CAF in use, which often leads to problems in the interpretation of research outcomes (Housen & Kuiken 2009: 462).

3.1 Defining Complexity, Accuracy and Fluency

Accuracy

To begin with, *accuracy*, also known as *correctness*, is empirically defined as "the extent to which an L2 learner's performance (and the L2 system that underlies this performance) deviates from a norm" (Housen et al. 2012: 4). The norm is usually the level of the native speaker and a deviation from the norm is termed error. One general problem arising from the definition of accuracy and its empirical application to learner language is: what target variety should serve as a norm and how strictly are deviations from the defined norm treated? Researchers suggest that *appropriateness* and *acceptability* in relation to a target variety should be taken into consideration when setting up marking criteria for accuracy measures. This is particularly important when considering non-standard or non-native usages that are fully acceptable in certain contexts (Housen et al. 2012: 4; for examples see Polio & Shea 2014: 22). Approaching accuracy in this way, however, should be carefully considered, as it might end in giving errors different weight according to their communicative effectiveness or putting them into relation to the developmental stage of learners. This might lead to measuring not only accuracy, but also other constructs such as development (Pallotti 2009: 592). Polio and Shea deem this point made by Pallotti as debatable, but they do acknowledge that considering the severity of an error does not raise reliability (2014: 22). They argue further that consent among groups of language users can be established, despite dialect and register issues (11). Furthermore, decisions that are taken by raters about what counts as an error should be explained and reported together with reliability measures that allow the replication of studies (Polio 1997: 129).

In essence, there are solutions to minimize subjectivity in the application of accuracy measures. Nevertheless, the subjectivity of the measure seems to lie in the general problem that the target norm itself may vary depending on context, language user or rater. For this reason, some researchers criticize that accuracy only measures the L2 knowledge of the learner as a proportion of the target norm instead of analyzing the L2 system of the learner itself. This has been referred to as “the comparative fallacy” (Bley-Vroman 1983; Thomas 1994: 328). However, it should be kept in mind that accuracy does not serve as a direct indicator of L2 development (Pallotti 2009: 592) and its purpose can only be seen as “the comparison with target-like use” (Wolfe-Quintero, Inagaki, & Kim 1998: 33).

At the same time, other researchers claim that accuracy is linked to the degree of correspondence of the L2 knowledge in the mind of the learner to the target norm in form of internalized rules and its degree of correct implementation in L2 production according to the rules of a target norm (Housen et al. 2012: 5; Skehan 1996: 46). Furthermore, other non-native L2 learners or the same learner at other stages of his or her L2 development can figure as the target norm to which accuracy is being related (Towell 2012: 52). Notwithstanding the debatable validity to directly measure the L2 system of language learners and the ongoing debates of error-definition for which some solutions seem to exist, accuracy is recognized as a well-defined, consistent construct (Housen et al. 2012: 4; Pallotti 2009: 592).

Fluency

The second component of the CAF triad to be defined is fluency, which has been long used for the general description of the learner’s language proficiency in relation to native-like spoken or written production. Nevertheless, *fluency* also refers to “the processing of language in real time” across modalities, with “speed and ease of processing” being components of it (Schmidt 1991: 92). Particularly, in L2 speaking it has been defined as “the production of language in real time without undue pausing or hesitation” (Ellis & Barkhuizen 2005: 139).

While the fluency construct for the modality of speaking seems empirically well defined (Skehan 2003: 8, 2009: 512–513), fluency in writing is argued to be determined by “the number of words or structural units” produced within a defined time frame (Wolfe-Quintero et al. 1998: 14). In contrast to this, other writing researchers suggests to tap fluency with language bursts, which are parts of new language generated in the writer’s mind and accessed through think-aloud protocols. This grounds on the assumption that language bursts strongly relate to the formulation process of Levelt's model. Hence, longer bursts indicate a higher degree of automatization (Chenoweth & Hayes 2001: 94).

In any case, empirical definitions of fluency are linked in some way to a temporal dimension. Therefore fluency is also described as the degree of automatization of proceduralized knowledge that allows the learner to access and retrieve available L2 knowledge for production. Consequently, fluent writing can be regarded as "automatic procedural skill […] not requiring much effort or attention" (Schmidt 1991: 93). Nevertheless, it should not be confounded with speed, because production speed could also result from fast processing which still consumes attention and effort; thus it is not automatized (Dörnyei 2009: 287). In conclusion, the construct of fluency should be able to depict the degree of efficiency with which learners are able to access and retrieve their productive L2 knowledge during writing (Wolfe-Quintero et al. 1998: 14).

Complexity

Defining *complexity* as component of CAF often leads to confusion. This results from the multiple usages in the context of different theoretical constructs, which is due to the polysemous nature of the term (Pallotti 2009: 593). Furthermore, an unclear distinction between related constructs has led to contradicting definitions (Bulté & Housen 2012: 22). To avoid this, the definition of complexity will be fully elaborated in chapter 4. This section will only briefly define the most important parts of the complexity construct to allow the explanation of the succeeding theories.

From a broad perspective, all terms describing aspects of complexity in language can be assigned to the two main approaches of *relative complexity* and *absolute complexity*[2]. The relative approach determines complexity on behalf of the mental ease with which the L2 user processes, learns or acquires a language item (Miestamo 2008: 25). Therefore, it is often equated with the term *difficulty* and describes complexity in relation to an individual language user (Wouter 2008: 9). In this way, the relative approach propagates a view that complexity in terms of acquisition and difficulty may vary individually between each learner and the learned language, including the L1 (10). As such, relative complexity is argued to describe language reality from the inside and comprises sub-components such as *developmental complexity* which is “the order in which linguistic structures emerge and are mastered in second (and, possibly, first) language acquisition” (Pallotti 2015: 118; Wouter 2008: 4). Another component associated with the relative approach is *cognitive complexity* i.e. the “relative difficulty with which language elements are processed during L2 performance and L2 learning, as determined in part by the learners’ individual backgrounds” (Housen et al. 2012: 4). This means that cognitive complexity shows learner dependent components that can be related to individual cognitive processing demands, abilities and perceptions. Consequently, these individual differences in the characterization of relative notions of complexity, always pose a tension on finding general, user-type-neutral criteria for its description (Miestamo 2008: 26).

---

[2] Some researchers criticize this term, because it suggests a theory-free, objective description of complexity. It is, however, in itself relatively bound to a certain linguistic theory. Nevertheless, *absolute complexity* will be used to remain consistent and avoid confusion with the main sources of this thesis (see Pallotti 2015: 119; Wouter 2008: 8).

In contrast to this, the *absolute approach* takes a structural and widely objective perspective on language complexity (24). This is also the approach that is pursued in this thesis. Therefore, complexity[3] can be defined as

> “the *number* of discrete components that a language feature or language system consists of, and as the *number* of connections between the different components” [original emphasis] (Bulté & Housen 2012: 24).

Following this definition, complexity can be described empirically by structural and intrinsic properties of a language that surface in production as quantifiable aspects of lexis, syntax, morphology or phonology. Consequently, *linguistic complexity* as a separate construct is entailed by the absolute complexity of a language, because it uses these linguistic properties of the language for its description. Syntactic complexity is one aspect of linguistic complexity and can be measured through its structural and formal components (26).

From a cognitive linguistic perspective, complexity in SLA is assumed to be associated with the degree of internalized L2 knowledge, consisting of procedures and linguistic items stored in the long-term memory of the learner that leads to a higher degree of variety and sophistication in expressing ideas during production (Housen et al. 2012: 5). As pointed out above, this link between linguistic complexity expressed in production and the actual amount of internalized L2 knowledge by the learner is rather inferred than empirically proven, because it is not possible to gain direct insight into the cognitive processes involved (Towell 2012: 56).

---

[3] The term *complexity* will be used henceforth in this sense, unless otherwise specified or modified by another word in front of the term which characterizes a different construct.

3.2 Trade-off hypothesis

Task-based language teaching (TBLT) research sees CAF as a reflection of L2 knowledge expressed in the learner's performance. From this perspective, TBLT aims to explain how certain language task variables affect CAF. In the last decades the outcome of this research field has not only shown that CAF are distinct performance areas (Skehan & Foster 1997: 204), but also lead to two controversial predictions about how task variables influence the interaction between these three dimensions (Robinson 2011a: 15; Skehan 2009: 527f.).

The first of the two theoretical assumption to be discussed is the Trade-off Hypothesis. It is also known as Limited Attention Capacity Hypothesis, which has been proposed by Skehan (1996, 1998). TBLT research associated with the Trade-off approach is based on a 'framework for task based instruction' to organize task features in conformity with the three variables of task type, task characteristics and task difficulty. *Task difficulty* is described in this framework solely as an objective, task inherent features influencing cognitive demands of a task, not accounting for individual differences in learners. The measurement of task difficulty still poses a problem in the description of tasks and has been operationalized differently within the field of TBLT (Skehan 2014b: 6). The three variables of task type, task characteristics and task difficulty are then related to the L2 task performance of learners, with CAF as dependent variables in the task outcomes.

The underlying main hypothesis of this TBLT research approach is that the cognitive resources of working memory and attention are limited. These limitations lead to trade-offs between the three performance areas reflected through CAF. Accordingly, learners are only able to direct attention to one of the three areas of CAF, while performance in the other two areas drops. Particularly, within the Trade-off Hypothesis it is assumed that, initially, attention can only be directed to either form or meaning. If the learner prioritizes meaning over form, this results in greater fluency. As a result, form, reflected in accuracy and complexity, is neglected. Furthermore, within the domain of form, a learner can direct attention to use challenging and new language features, resulting in more complex performance, or to focus attention towards already known language and seeking greater control over production, leading to greater accuracy (Skehan 1998: 179, 2009: 511).

Following the Trade-off Hypothesis, it has been argued that tasks can be designed in a way that they provide opportunities for learners to choose one of the three performance areas over the other; thus, manipulating and extending their L2 system accordingly. For instance, tasks that are implemented in a way to let the learner focus on aspects of form such as the need for appropriate and correct language production or the use of language at the upper limit of their L2 system, are likely to raise accuracy or complexity respectively. Although task characteristics might enhance the development of a certain area through pushing attention towards one dimension of CAF, the influence of individual differences seem to be non-negligible factors in task-design. While, for instance, a learner might realize through task conditions that the own L2 system is limited, some learners are more willing to use this opportunity and try new L2 structures, resulting in more complex, but less accurate language, than others (Skehan & Foster 2001: 190f).

More important for the purpose of this thesis is how the trade-off approach links task variables and performance outcomes theoretically to Levelt's model of speech production. The research results from TBLT contribute to the validation of a theoretical, cognitive model for L2 production (Skehan 2009: 524). In this regard, the trade-off approach identifies the conceptualizer stage in Levelt's model as the processing component mainly responsible for the complexification of L2 production. For this reason, task properties requiring the formulation of complex ideas which need cognitively demanding organization and integration of information, result in higher structural complexity (520). If a task provides more structure, on the contrary, it leads to less pressure on the conceptualizer and frees attention for the subsequent formulator stage. Consequently, more available attention for the formulator stage is claimed to have positive effects on fluency and accuracy in L2 production (519).

3.3 Cognition hypothesis

While the described trade-off approach to TBLT research is concerned with how task design can be influenced to change the outcome of one or more of the three CAF variables, another perspective is taken by research conducted within the Cognition Hypothesis, as proposed by Robinson (2001, 2011a). Research related to the Cognition Hypothesis uses the Triadic Componential Framework for task characterization (Robinson 2011a: 6). This framework allows the classification of tasks according to sequences of progression over periods of instruction, according to their cognitive complexity (Robinson 2003: 61). In contrast to Skehan's framework for task based instruction, Robinson uses the terms *task complexity* and *task difficulty.* Task complexity accounts for the cognitive demands of L2 tasks that are objectively controllable, while task difficulty describes the cognitive demands based on differences and factors determined by the individual learner (Robinson 2001: 295). Within the Triadic Componential Framework, however, a distinction between resource-directing and resource-dispersing task dimensions is made at the level of task complexity. *Resource-directing* task variables impose conceptual or cognitive task demands on the learner such as spatial reasoning or the number of task elements. These variables are able to point the learner towards language features such as logical subordinators or deictic expressions which can be used to solve L2 tasks. According to Robinson, a rising cognitive demand in resource-directing variables enhances the mapping of new concepts and functions of forms in the L2 (Robinson 2003: 58). In contrast to this, resource-dispersing task characteristics require performative and procedural demands from the learner such as planning time or prior knowledge and might promote real-time access to existing L2 knowledge (59).

The Cognition Hypothesis predicts that sequentially raising task complexity along resource-dispersing task dimensions will have a negative impact on fluency, accuracy and complexity in L2 production. On the contrary, sequencing resource-directing task characteristics from simple to complex promotes greater complexity and accuracy through connecting cognitive resources and demanding greater effort in the conceptualization of the L2 message (Robinson 2011a: 14). In other words, processing conceptually more demanding ideas for L2 production leads not only to higher cognitive demands, but also simultaneously facilitates the accuracy and complexity of L2 performance (Robinson 2011a: 15). Although the Cognition Hypothesis makes no direct claim about fluency, it is proposed that raising task demands on both resource-dispersing and resource-directing dimensions might have positive effects on general L2 production (Robinson & Gilabert 2007: 167).

This is also reflected in the underlying main principle of the Cognition Hypothesis motivated from L1 research, but adapted for L2 purposes. In essences, it states that complex conceptualizations of the world, as they are perceived by each individual speaker, require complex language to express those concepts (Robinson 2011a: 14). Consequently, language tasks should help the adult learner to gradually build up the L2 inventory for expressing already known, complex concepts from the L1. The rising demands on conceptualization along resource-directing task variables are also claimed to be the main factor that links the Cognition Hypothesis to Levelt's model of speech production. Robinson argues that the higher conceptual demands from language tasks result in greater demands on macro- and micro-planning at the conceptualizer stage for the preverbal message. In this way, L2 tasks promote new ways of thinking which are required for effective L2 production and enhance the mapping of new procedures in the mind for transforming concepts form the L1 into linguistic expressions of the L2 (15). The so formed preverbal message requires more complex encoding processes. Consequently, high conceptual task demands trigger the extension of the L2 conceptual system and in this way syntactic development (Kormos 2011: 54).

L2 complexity mainly develops through processes at the conceptualizer stage, as proposed by the above linkage of the Cognition Hypothesis to Levelt's model. This is similar to the Trade-off Hypothesis. Additionally, both models propose the sequencing of L2 tasks according to their cognitive demands, although each approach operationalizes these demands in other ways. The differences in the operationalization of concepts such as task difficulty or task complexity often leads to the use of different CAF measures and seems to make the verification of each hypothesis difficult (Housen et al. 2012: 6; Robinson 2011a: 21; Skehan 2014a: 241). Despite their controversial claims, each of the two hypotheses delivers not only valuable contributions for TBLT, but also contributes to the description of the interrelations between CAF. Therefore, the assumptions made by each model can be seen as one perspective of how to look at CAF in L2 production.

3.4 Dynamic System Theory

In contrast to the hypotheses from TBLT research which adhere to cognitive models of SLA for explaining their study outcomes, Dynamic System Theory (DST) presents a different approach to the interpretation of L2 data. DST as a transdisciplinary research framework allows to interpret CAF with a focus on L2 development and change in combination with other approaches to SLA (Schmid, Verspoor, & MacWhinney 2011: 26; Thelen & Smith 1994). Researchers following the DST approach are primarily interested in how L2 change and variability in individual learners can be depicted over time, while not assuming a theoretically finite state of development. Furthermore, DST understands language learning as a continuous, never ending sociocognitive process that is influenced by the environment (Cameron & Larsen-Freeman 2007: 141). According to DST theory, language is expected to be a non-linear, complex and dynamic system, consisting of interdependent sub-systems that adapt and develop in various ways (Larsen-Freeman 1997: 149). Due to its transdiciplinarity, DST acknowledges a wide range of sources for evidence. Its view on language development, however, is expected to be partially reflected in the multidimensionality of CAF and it is suggested to operationalize a range of CAF sub-constructs to get a preferably wide picture of their development at different levels of measurement granularity (Larsen-Freeman 2015: 232; Schmid et al. 2011: 43).

Interestingly, with regards to the before mentioned trade-off approach to TBLT, DST also assumes trade-offs in certain areas of development. This explains the wide range of variables used to tap different areas of development, because improvement in one area, for instance lexical variation, might result in a decline in other areas such as syntactic complexity (40). DST assumes that various areas of L2 development are interconnected and are able to be competitive, supportive or conditional in complex ways (Larsen-Freeman 2006: 592f). Moreover, it is the main intention of DST to describe these non-linearities and processes of change through small-stepped longitudinal studies of individual learners that should help to explain these interconnections and possibly allow the mathematical modeling of such interactions (Cameron & Larsen-Freeman 2007: 16; Caspi 2010; Lowie, Caspi, Van Geert, & Steenbeek 2011: 99f). Consequently, this research approach goes in line with suggestions for a more organic, multivariate and longitudinal account for the general investigation of CAF constructs, measures and their interconnections to provide a better understanding of the same (Norris & Ortega 2009: 574).

The theoretically implied use of CAF to trace L2 development leads up to a general problem: DST still lacks suitable measures to allow a definition of L2 development on behalf of CAF (Larsen-Freeman 2009: 580). The missing construct definition for L2 development expressed through CAF in some DST studies, is criticized by other researchers (Pienemann 2007: 44). Moreover, the non-linearity and unpredictability assumption of L2 development as proposed by DST theory has been claimed not to be true (Baten & Håkansson 2015: 541). At the same time, it has been proposed that CAF measures might not be suitable to measure L2 development (520). These claims, however, come from Processability Theory research; a theory that defines development according to sequences of linguistic forms that can be handled at a certain stage of development by a learner (Pienemann & Lenzing 2015: 159). While problems with CAF as measures for L2 development have already been mentioned before (see 3.1), it has been argued that part of the criticized points result from theory-inherent differences of both approaches (Lowie & Verspoor 2015: 81; Verspoor, Lowie, & Van Dijk 2008: 216).

Apart from the controversial aspects of DST and its relative young nature as a research approach, its transdisciplinarity promises potential to explain interconnections in CAF and L2 development (Norris & Ortega 2009: 556). Moreover, the reliance on mathematical modeling could be of help to provide an alternative view to L2 development and provide valuable contributions for language pedagogy (Lowie et al. 2011: 121; Weideman 2010: 232).

4 Defining Complexity

4.1 L2 complexity

This sub-chapter focuses on the explanation of individual parts of L2 complexity and points out the place syntactic complexity takes within this construct. Most of the sub-components comprised by L2 complexity consist of several other constructs. A rough outline of them can be seen in the taxonomy of L2 complexity constructs provided by Bulté and Housen in figure 2 (2012: 23).

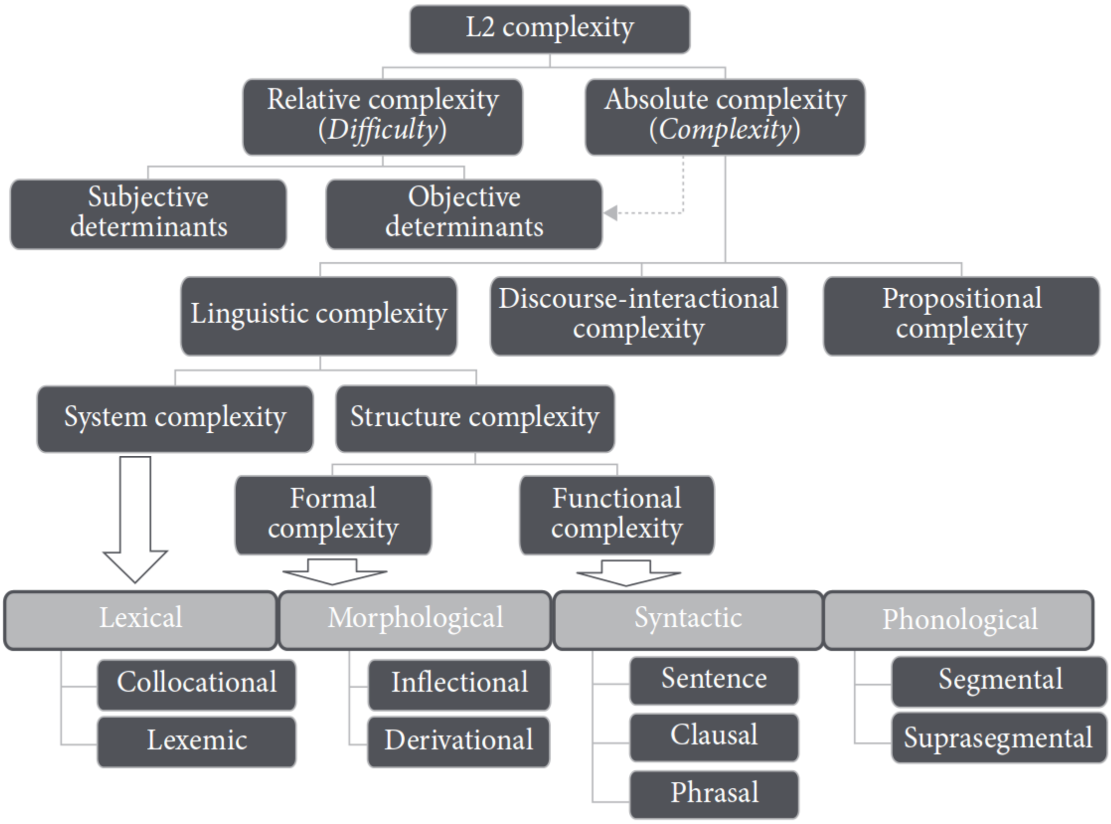

L2 complexity
Relative complexity (*Difficulty*)
Absolute complexity (*Complexity*)
Subjective determinants
Objective determinants
Linguistic complexity
Discourse-interactional complexity
Propositional complexity
System complexity
Structure complexity
Formal complexity
Functional complexity
Lexical
Morphological
Syntactic
Phonological
Collocational
Lexemic
Inflectional
Derivational
Sentence
Clausal
Phrasal
Segmental
Suprasegmental

Figure 2: A taxonomy of complexity constructs (Bulté & Housen 2012: 23)

L2 complexity can be broken down into the two major approaches of *relative complexity* and *absolute complexity* to describe complexity in language, as shown in figure 2 and as already mentioned before (see 3.1). Relative complexity or (cognitive) difficulty, can be further distinguished into *subjective determinants* that are solely dependent on the learner such as memory capacity, motivation or stage of L2 development and *objective determinants*. The latter include language features that have been demonstrated to be more difficult to process or acquired at a later stage in language development (e.g. embedded relative clauses) than other language features (e.g. coordinated clauses) (Bulté & Housen 2012: 23). At the same time, objective determinants are also described by the absolute approach such as the amount and variety of subordination or the number of propositions included in a text. This relation is indicated by the dotted line in figure 2. This dotted line also represents the theoretical main link that connects the objective description of complexity with the theoretical cognitive and learner dependent accounts. Accordingly, this relationship assumes that "[m]ore complex mentally-represented events are coded by more complex linguistic/syntactic structures"; hence they demand higher cognitive effort in processing. Consequently, syntactically more complex structures result in "more complex mental processing operations" (Givón 2009: 12). However, the possible correlation of linguistic complexity and difficulty is only inferred, but has not been validated (12). Despite this intuitively and logically correct assumption, a one-to-one relationship between the two constructs of difficulty and complexity cannot be taken for granted and is still subject of scientific investigation (Bulté & Housen 2012: 24).

The analysis of learner language in the methodological part of this thesis, however, will concentrate on describing complexity from an absolute perspective. The basic idea of the *absolute approach* is to describe complexity by the number of its discrete language elements and the connections between them in objective and quantitative ways as proposed by the definition of complexity in chapter 3.1 (23). The construct of absolute complexity can be further distinguished into *propositional complexity*, *discourse-interactional complexity* and *linguistic complexity*. On the one hand, propositional complexity is claimed to be expressed in a production by the amount of information that is expressed by a text measured in, for instance, idea units. Discourse-interactional complexity, on the other hand, has not been fully investigated so far (24). Attempts to describe discourse-interactional complexity usually include the description of number and type of turn changes or the frequency of cohesive devices and their implementation according to spatial, temporal, causal or intentional discourse relations (Kormos 2011: 203). Although all three sub-constructs are equally important in investigating L2 complexity, linguistic complexity is of major importance for this thesis.

According to Bulté and Housen's taxonomy shown in figure 2, *linguistic complexity* comprises the two sub-constructs of system complexity and structure complexity. S*ystem complexity* is reflected in breadth, range and richness that a piece of writing expresses in relation to the number and range of structures and linguistics items which are available in the language system under investigation. System complexity roughly corresponds to global complexity features of a language system which, hypothetically, can be seen as the number of elements and structures a whole language system consists of. In contrast to this, *structure complexity* is associated with depth of individual linguistic features at a local level and is reflected in the embedding and composition of forms, patterns and structures that make up the subsystems and layers of a language system (Bulté & Housen 2012: 25, 2014: 44). Within the structural make-up of a L2 system, a further distinction can be made between formal complexity and functional complexity.

Formal complexity refers to the linguistic entities that are needed to build a certain linguistic feature. An example would be the construction of the simple past in contrast to the present perfect form, which takes more words to be realized. In this sense, it is associated with the number of operations to arrive at a target structure when starting from a base structure, but also with the distance between a form and its dependents. Different from this is *functional complexity*, which describes the instances of form to meaning or function mappings that can be expressed by a linguistic entity. It accounts for the transparency or multiplicity of meanings and functions that can be applied to an linguistic item. For instance, the English plural marker '-s' directly maps the meaning of plural onto a noun, while the form of the word 'present' can figure as a verb or a noun and thus expresses more meanings accordingly (Bulté & Housen 2012: 25).

As this outline reveals, the theoretical construct of L2 complexity and its comprised sub-components are multi-dimensional. Although they are closely related and intertwined in reality, in theory these sub-constructs are all distinct from each other. This makes the separate assessment of them difficult. Nevertheless, the sub-constructs of system, formal and functional complexity surface in the different linguistic domains of lexis, morphology, phonology and syntax, as indicated by the three white arrows in figure 2. As such, they are expressed in L2 production and can be evaluated accordingly. For instance, such an evaluation then allows to form conclusion about the system complexity of the syntactic L2 system of a learner or about the functional and formal complexity of the learner's syntactic features that have been applied in his or her writing (26). It should be noted that each of the language domains is only able to depict one aspect of L2 complexity in the L2 system investigated. Moreover, each language domain in itself has again different dimensions, as indicated by the boxes below them in figure 2. In the case of syntax these dimensions are argued to be the sentence, clausal and phrasal level (23).

4.2 Syntactic complexity

This part briefly discusses the various ways of how L2 complexity manifests itself in the domain of syntax and with which means to measure it. In order to accomplish this, the three construct levels of theory, observation and operation will be specified following Bachman's suggestion for language assessment (1990: 40–45). For a better understanding of the different construct levels, this specification will take advantage of Bulté and Housen's visual representation of the theoretical construct of *grammatical complexity*, as shown in figure *3*. Grammatical complexity is one major sub-component of linguistic complexity (Bulté & Housen 2012: 26). Although grammatical complexity comprises the language aspects of syntax and morphology, as shown at the observational level in figure 3, the specification presented here will only concentrate on syntax.

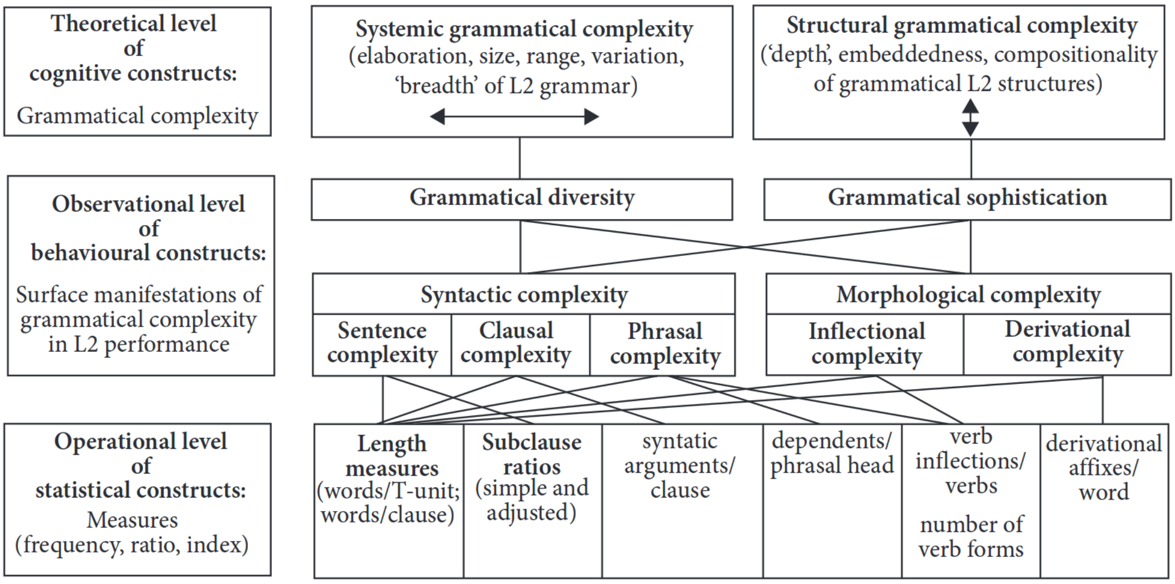

Theoretical level of cognitive constructs:
Grammatical complexity
Systemic grammatical complexity (elaboration, size, range, variation, 'breadth' of L2 grammar)
Structural grammatical complexity ('depth', embeddedness, compositionality of grammatical L2 structures)
Observational level of behavioural constructs:
Surface manifestations of grammatical complexity in L2 performance
Grammatical diversity
Grammatical sophistication
Syntactic complexity
Sentence complexity
Clausal complexity
Phrasal complexity
Morphological complexity
Inflectional complexity
Derivational complexity
Operational level of statistical constructs:
Measures (frequency, ratio, index)
Length measures (words/T-unit; words/clause)
Subclause ratios (simple and adjusted)
syntatic arguments/ clause
dependents/ phrasal head
verb inflections/ verbs
number of verb forms
derivational affixes/ word

Figure 3: Construct definition of grammatical complexity (Bulté & Housen 2012: 27)

At the *theoretical level*, as depicted at top of figure 3, syntactic complexity is able to account for aspects of system complexity and structure complexity as cognitive manifestations of grammatical complexity in the learner's mind. S*ystem grammatical complexity* corresponds to the breadth, range and size of the (cognitive) L2 grammatical system under investigation, while *structural grammatical complexity* refers to the the composition of forms, their integration and the relations that exist between them in the mind of the learner (26). These cognitive aspects of the L2 system are then expressed by the learner in L2 production.

A concrete example of such a structure at the sentence level would be the realization of conjoining clauses by using coordination or subordination to achieve certain syntactic functions. At the clausal level, for instance, syntactic complexity can be expressed by the number or type of syntactic arguments that are used in a clause to express meaning. An example of phrasal complexity would be the number or type of noun phrase dependents a learner uses to realize such a phrase. Nevertheless, such structures are part of the *behavioral construct* and can be observed in language samples through diversity and sophistication of syntactic features. These syntactic components can then be assessed with different measures. They constitute the *statistical construct* of syntactic complexity, in other words how syntactic complexity is operationalized to analyze L2 writing. There are different ways to measure syntactic complexity, as the statistical construct level in figure 3 indicates. Some measures simultaneously account for complexity in all three dimensions (e.g. words per T-unit), while others are more specific and only tap one dimension (e.g. syntactic arguments per clause only accounts for clausal complexity) (27). This is indicated by the lines between the behavioral and the statistical construct level in figure 3.

In conclusion, this means that each measure yields different values in terms of meaningfulness for different sub-constructs and allows different conclusions at the theoretical level about the L2 system under investigation. Each research tradition, therefore, uses different measures and conceptualizations of syntactic complexity. In traditional linguistics, for instance, syntactic complexity is mainly treated as a phenomenon at the sentence and clausal level and assessed through number and type of coordinations and subordinations. Another example are functional approaches to text and genre analysis that emphasize the operationalization of syntactic complexity at the phrasal level. Moreover, the functional approach puts more weight on the variety of syntactic structures and therefore it is accompanied by labor intensive assessment to arrive at the indices needed for such an analysis. This thesis takes on an intermediate position and applies more economic measures, but at the same time accounts for syntactic complexity at the clausal and phrasal level to some extent. Such an approach has also been propagated by recent studies to assess L2 complexity in learner language (Biber et al. 2016: 649; Yang et al. 2015: 54). The following sub-chapter presents such an example of a construct definition for syntactic complexity. This definition is partly used in the methodological part of this thesis.

4.3 Operationalizing syntactic complexity

Despite the vast amount of different measures and operationalizations in L2 literature, Yang, Lu and Weigle present a construct definition that efficiently accounts for the multiple dimensions of syntactic complexity in academic L2 writing (2015: 55). This allows the comparison of results from other studies and contributes to the validation of these measures (Norris & Ortega 2009: 574). Moreover, the variety of measures seems appropriate to cover a range of developmental tendencies (568). For this reason, an adaptation of their construct operationalization, as shown in figure 4, is used in this thesis for the analysis of the investigated data[4].

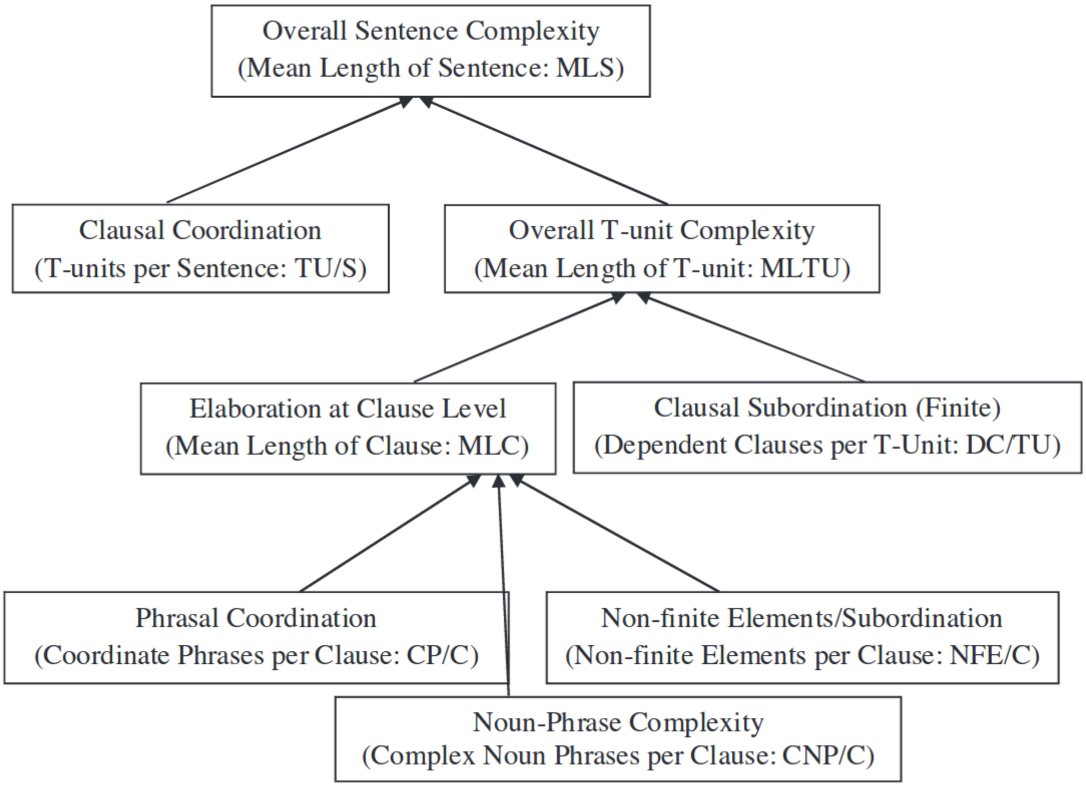

Overall Sentence Complexity
(Mean Length of Sentence: MLS)
Clausal Coordination
(T-units per Sentence: TU/S)
Overall T-unit Complexity
(Mean Length of T-unit: MLTU)
Elaboration at Clause Level
(Mean Length of Clause: MLC)
Clausal Subordination (Finite)
(Dependent Clauses per T-Unit: DC/TU)
Phrasal Coordination
(Coordinate Phrases per Clause: CP/C)
Non-finite Elements/Subordination
(Non-finite Elements per Clause: NFE/C)
Noun-Phrase Complexity
(Complex Noun Phrases per Clause: CNP/C)

Figure 4: A multidimensional representation of syntactic complexity (Yang, Lu & Weigle 2015: 54)

Each of the boxes in figure 4 represents a separate sub-construct of syntactic complexity that can be observed by the indices that are given in parenthesis. Furthermore, the arrows point out the hierarchical relationship between the different sub-constructs in terms of the syntactic level they are able to capture. While the non-terminal nodes (i.e. the sub-constructs at which arrows point at) represent more than one single construct or tap into more than one dimension of syntactic complexity, the terminal nodes (i.e. sub-constructs that point to a hierarchical higher sub-construct) represent discrete sub-constructs of syntactic complexity. The three composite or holistic sub-constructs which capture more than one distinct syntactic element are *overall sentence complexity*, *overall T-unit complexity* and *elaboration at clause level*. In contrast to these composite sub-constructs, there are the five distinct sub-constructs of *clausal coordination*, *finite clause subordination*, *phrasal coordination*, *non-finite elements per subordination* and *noun-phrase complexity* which tap only one specific syntactic element. The sub-construct of noun-phrase complexity is argued to be an important indicator of development in the genre of academic writing. On the one hand, this has to do with the fact that academic texts employ more noun phrases; therefore, noun phrases are considered to be a characteristic feature of academic texts (Crossley, Roscoe, & McNamara 2014: 204; Mazgutova & Kormos 2015: 7). On the other hand, it has been proposed by systemic functional linguistics that writers move from dynamic writing styles to more synoptic writing styles. In terms of syntax this means that information is expressed by beginner writers mainly through coordination and with increasing proficiency through interdependent clauses. Moreover, highly advanced learners condense the same information to the phrase or word level through phrasal complexification (Halliday & Martin 1993: 43–54; Norris & Ortega 2009: 563). Although the construct of noun phrase complexity is of special interest to the genre of academic writing, it was not possible to modify the analysis software accordingly[5] (for further information see Yang et al. 2015: 58–9).

---

[4] The original version of Yang, Lu and Weigle's construct definition includes a measure for noun phrase complexity, as shown in figure 4. This, however, could not be realized with the methods at hand for this thesis; therefore this measure has not been utilized in the empirical part.

Sequencing units

Before each of these sub-constructs can be discussed in more detail, it is necessary to define the units in which language samples are sequenced to account for different instances of syntactic complexity. The following definitions are used in this form also in the methodological part and all of them are in line with the measures used by Yang, Lu and Weigle (2015: 58). To begin with, a *sentence* is interpreted as “a group of words delimited with [a] period, exclamation mark or question mark” (Lu 2010: 481). (1), (2), (3) and (4) are all examples for sentences and their length in words is indicated at the end of each example in curly brackets[6].

---

[5] It has been assumed that the software modifications could be easily accomplished, but it would have consumed too much time and effort for the scope of this thesis. It has also been considered to use a measure for complex nominals which is provided by the analysis software instead, but its definition proved to be too time consuming in the validation procedure. Nevertheless, the index of complex nominals has been successfully used in other studies (see Lu 2010).

(1) **It is questionable** /$_{3}$ whether Lincoln would have gotten the status of a hero /$_{10}$ if he had been forced to speak on national television^{5}, /$_{10}$ |^{23} **but still, his story reveals the importance of charisma in political language of democratic nations**. /$_{15}$ |^{15} {38}

(2) **Herdina and Jessner argue** /$_{4}$ that monolinguals, bilinguals and multilinguals employ different strategies when learning a foreign language^{5}. /$_{13}$ |^{17} {17}

(3) **The last part consists of pair associations** /$_{7}$ that have to be memorized.^{3} /$_{5}$ |^{12} {12}

(4) **In order to share and spread these oppositional ideas**^{9}**, political parties need to make use of language as a means of expression and persuasion**^{12}. /$_{24}$ |^{24} {24}

Defining the sequencing unit of a *clause* is already a more complicated undertaking. Although general linguistic theories count structures that contain a finite or a non-finite verb as separate clauses, most definitions of a clause in writing research, including this thesis, follow Hunt's definition of a clause. He defines a clause as a "structure with a subject and a finite verb (a verb with a tense marker)" (Hunt 1965: 15). This definition of a clause therefore includes independent clauses, nominal clauses, adjective clauses and adverbial clauses (Lu 2010: 481). Consequently, the sentence in (1) consists of four separate clauses, while (2) and (3) are only composed of two clauses per sentence and sentence (4) is interpreted as one clause. Clauses are indicated by a slash at the end of each clause. The numbers in subscript indicate the clause length in words. A clause can then be further distinguished into a *main clause* that can stand alone as a separate sentence or a *dependent clause* that is attached to a main clause and can be characterized as a finite nominal, adjective or adverbial clause (482). Example (1) comprises two main clauses, which are marked in bold print, and two finite dependent adverbial clauses that are underlined, while (2) shows an example of a dependent finite nominal clause. The dependent clause in (3) represents an example of a dependent finite adverbial clause. All dependent clauses are marked by underlining in the given examples.

---

[6] The examples have been taken from text 28_SE, sentence number 66 ex.(1) and sentence number 23 ex.(4), from text 1_SE, sentence number 51 ex. (2) and from text 58_SE, sentence 36 ex.(3). All three texts can be found in the appendix with the according numbers of sequencing units annotated for each sentence.

Despite not counting as a separate clause in the approach pursued here, *non-finite elements* are still seen as instances of subordination and therefore they are defined as verb phrases headed by a non-finite verb (Lu 2010: 484; Yang et al. 2015: 59). Examples for such non-finite elements can be found in (1) and (3) which are headed by the to-infinitives *to speak* and *to be*. Example (2) shows one non-finite construction of five words that is headed by the infinitive *learning*. Non-finite constructions are marked by gray highlighting and their length is indicated in number of words written in superscript at the end. Another sequencing unit that is applied at a sub-clausal level is the coordinate phrase. *Coordinate phrases* are all adjective, adverb, noun and verb phrases "that immediately dominate a coordinating conjunction" (Lu 2010: 483). Two instances of coordinate noun phrases can be found in (2) and one coordinated verb phrase and one coordinated noun phrase are part of (4). The coordinated phrases are indicated with overlining.

The last sequencing unit that is very specific to writing research is the T-unit. It has been introduced by Hunt to avoid instances of endless coordination in beginner writing to be assessed as complex structures. Furthermore, the T-unit allows to assess subordination separately from coordination (Hunt 1965: 21). A *T-unit* can be defined as "one main clause plus any subordinate or non-clausal structure that is attached to or embedded in it" (Hunt 1970: 200). Following this definition, (1) can be subdivided into two T-units, while (2), (3) and (4) consist only of one single T-unit. T-units are marked by a vertical line after each T-unit and the numbers in superscript indicate the length of the T-unit in words. It can be seen in (1) that the coordination realized with *but* between the two T-units cannot be detected if the sample is only examined at the level of T-units. Instances of coordination are therefore assessed through a separate measure to account for instances of coordination at the sentence level.

Global measures

After this short outline of sequencing units, which are necessary to operationalize the sub-constructs of syntactic complexity (see figure 4), it will be explained how exactly these sub-constructs are operationalized. To begin with, the two composite constructs of *overall sentence complexity* and *overall T-unit complexity* are operationalized through the length measures *mean length of sentence* (MLS) and *mean length of T-unit* (MLTU) respectively. Such measures are a convenient way to capture global syntactic complexity by assessing the mean length of words per analytical unit – in this case sentence or T-unit. The underlying main assumption is that an increase of words per analytical unit is a reflection of higher syntactic complexity, because this increase can only happen through elaboration at levels that are below that of a sentence or T-unit (Yang et al. 2015: 55). For instance, the sentence in (1) comprises 38 words, and is therefore considered to be more complex than the sentence in (4) consisting of 24 words. However, when calculated in MLTU each of the two T-units in (1) count less words than the T-unit in (4); thus both T-units in (1) are considered less complex compared to the T-unit in (4). The difference that can be seen between the two operationalizations is that the MLS allocates the same weight to coordinated and subordinated clauses in a sentence, while the MLTU does not account for compound sentences (Bulté & Housen 2014: 49).

A closer look at the two examples above reveals the weakness of length based measures. For instance, the first T-unit in (1) has two subordinate clauses, which can be interpreted as another form of syntactic complexity, whereas the T-unit in (4) has more words, but no clausal subordination. This illustrates that such global measures do not indicate at which level or in which form syntactic complexification happens. In other words, the MLS and the MLTU do not account for hierarchically lower sub-constructs in a discrete way and some researchers argue that they confound the assessment of discrete syntactic functions. Moreover, they are argued to lack a theoretical linguistic basis (Biber et al. 2016: 649; Ortega 2003: 493). Nevertheless, the two measures tap different levels and various sources of syntactic complexity simultaneously. For this reason, they are considered more practical when representing an overall picture of syntactic complexity compared to more specific measures (Bulté & Housen 2012: 36). Moreover, they have been shown to correlate at least to some extent with the linguistic reality of syntactic complexity in learner language (Ortega 2003: 512; Wolfe-Quintero et al. 1998).

Local measures

Another more specific length-based measure of the construct outlined in figure 4 is based on the clause as unit of analysis. It represents the composite sub-construct of *elaboration at clausal level*. In contrast to global measures, the *mean length of clause* (MLC) is considered to be more specific and accounts for syntactic complexity at the subsentential level (i.e. clausal and phrasal complexity) (Norris & Ortega 2009: 561). The MLC is defined by the number of words per clause. This measure is based on the assumption that the number of phrases in a clause is limited. As a consequence, phrasal elaboration results in an increase of clause length. However, it should be noted that an increase of clause length can also arise from adding other clause constituents such as adjuncts (Bulté & Housen 2012: 38). Another issue of the MLC for representing clausal and phrasal elaboration is that it depends crucially on how a clause is defined and operationalized. This has to be taken into account when interpreting study results or if comparing results with other studies. For instance, considering the non-finite element of *to speak on national television* in the third clause of (1) as a separate non-finite clause, would result in a higher overall clause count and decrease the MLC. The same is true for (2) and (3) which would consist of three clauses instead of two in this case and (4) would yield a clause count of three instead of one. Non-finite constructions, however, are considered to be part of a clause in the definition that is adhered here and only represent one specific form of verb phrases. Although this conceals the exact place of complexification, the practical application of the measure and its empirical validity to tap expected sources of complexity at the phrasal level speak for the application of the MLC (Biber et al. 2016: 649).

Two other constructs that also account for syntactic complexity at a local level are the sub-constructs of *clausal coordination* and *phrasal coordination*. In contrast to the sub-constructs discussed so far, each of the following sub-constructs represents one specific instance of syntactic complexity. According to figure 4, coordination at the clausal level is operationalized through the index of *T-unit per sentence* (TU/S). The two T-units in (1) that are coordinated by *but* demonstrate the empirical validity of this operationalization, because a T-unit comprises at least one main clause and clausal coordinations only appear between two main clauses. The example in (1) consists of a compound sentence composed of two T-units and therefore raises this index by one, whereas (2), (3) and (4) only constitute one sentence and one T-unit which keeps the index at one. Coordination at the phrasal level, however, is operationalized in a similar way, but taps the sub-clausal level through the construct of *phrasal coordination* which is realized as *coordinated phrases per clause* (CP/C). This sub-construct indicates one way of syntactic sophistication and elaboration at the phrasal level. Examples can be seen in each of the two coordinated phrases in (2) and (4) that both illustrate how phrasal elaboration through coordination is employed to raise the word count per clause (Yang et al. 2015: 55).

Another way to examine the syntactic means for clausal elaboration is indicated by the sub-construct of *non-finite elements per subordination.* This construct is represented through the number of *non-finite elements per clause* (NFE/C). Although non-finite structures are not counted as separate clauses according to the definition above, they are still viewed as an instance of subordination at the clausal level (54-55). The third clause in (1), for instance, shows how the non-finite element *to speak on national television* contributes to clause lengthening. Similarly, both non-finite constructions in (4) contribute essentially to the length of the main clause and illustrate how the NFE/C reveals this source of complexification. However, all of the above examples also point out that this operationalization is again crucially dependent on the definition of a clause, as this has been pointed out before. When adhering to the clause as defined in this thesis, the NFE/C, however, represents a form of non-clausal subordination.

The last construct to be discussed represents subordination at the sentence level. *Finite clausal subordination* accounts for syntactic elaboration at the sentence level through the number of *dependent clauses per T-unit* (DC/TU). The example in (1) shows how the two dependent clauses contribute to sentence or T-unit length, while in (4) there is no such instance and the length of the main clause is constructed by other means such as non-finite elements or compound phrases at the sub-clausal level. The above considerations about the clause definition also apply to this sub-construct in terms of other measurement outcomes.

The TU/S, the DC/TU, the CP/C and the NFE/C only account for specific instances of subordination or coordination in the corresponding dimension; thus they are relatively narrow in terms of their linguistic scope (Bulté & Housen 2012: 37). Therefore, it appears logical to adhere to recommendations for the application of a construct that assesses the multidimensional nature of syntactic complexity. The construct as a whole represents an attempt to accomplish this, while it provides a reasonable path between the amount of measures applied and the systematic and structural aspects displayed by the selected sub-constructs (Biber et al. 2016: 649). Considering this, it should always be kept in mind that "what is complex according to one theory may not be so according to another" (Pallotti 2015: 119).

4.4 Syntactic complexity in L2 academic writing

Syntactic complexity is investigated with respect to three major goals: to trace L2 development, to assess L2 proficiency and to describe language task outcomes (Ortega 2012: 128). As such, studies investigating syntactic complexity contribute to a better understanding of learning processes in the L2 and help in the development of pedagogical language tasks (see Housen et al. 2012; Larsen-Freeman 2006; Robinson 2011b; Skehan 2014b).

Based on developmental language patterns from systemic functional linguistics, in earlier studies it was proposed that overall syntactic complexity at beginner levels mainly unfolds through coordination. This is followed by a higher amount of subordination employed at intermediate proficiency levels. Finally, subordination is expected to be superseded by sub-clausal elaboration through phrasal modifications at highly advanced proficiency levels (Halliday & Martin 1993: 43–54; Norris & Ortega 2009: 562). Recent corpus studies of syntactic complexity at different proficiency levels, however, did not support all of these assumptions about syntactic development. Although the data was collected over a rather short developmental time span of four months, the results suggested that coordination of clauses regained importance at intermediate proficiency levels (Bulté & Housen 2014: 53). However, other studies support the assumption of elaboration at the phrasal level and suggest a higher variation in the use of different syntactic constructions with increasing L2 proficiency (Crossley & McNamara 2014: 74). Both findings are, however, in line with results from DST studies that syntactic complexity develops in unpredictable ways over short periods of time with increase described in certain dimensions, but decrease, stagnation or even backsliding in others (Spoelman & Verspoor 2010: 551). Notably, Norris and Ortega also refer to the DST approach to support their assumed developmental pattern outlined at the beginning of this paragraph (2009: 573).

Nevertheless, over a period of ten months, an increase in global measures of syntactic complexity, such as the MLTU, has been documented in five adult Chinese learners of English while receiving continuous instruction (Larsen-Freeman 2006: 604). Similarly, a recent study demonstrated that a modest rise in syntactic complexity could be seen in lower proficiency groups in academic genres already after a one-month period of intensive English for academic purposes (EAP) instruction. Interestingly, the authors of the same study stated that advanced L2 learners of English showed little improvement in syntactic development after general short-term EAP instruction (Mazgutova & Kormos 2015: 11–13). Similar results were found in an earlier study of foreign students in England who took three months long EAP courses to improve their already advanced academic writing (Shaw & Liu 1998: 244). However, Mazgutova and Kormos also found evidence for the previously outlined developmental pattern, namely that syntactic development in academic writing at higher proficiency levels was mainly achieved due to phrasal elaboration via complex nominals and noun phrase modification. Furthermore, they observed non-significant changes in these measures after a four months period of EAP instruction at this proficiency level (2015: 12).

L2 proficiency is often confused with syntactic development, because some indices suggest a relationship between these two distinct constructs. The fact that syntactic development is not equivalent to L2 proficiency has been demonstrated in a study that investigated 10 syntactic complexity measures (Bulté & Housen 2014: 47). The results showed that not all measures of syntactic complexity correlated with proficiency scores (56). Moreover, Bulté and Housen reported a significant correlation of only half of the syntactic complexity measures with holistic writing proficiency scores (52). Crossley and McNamara argued in the same way when comparing certain syntactic complexity features that indicated growth patterns with syntactic elements used by raters to judge proficiency (2014: 76). It should be mentioned, however, that Crossley and McNamara concentrated on phrasal elaboration and even considered syntactic transformations (e.g. negations or wh-questions) in their construct definition of syntactic complexity (70). Their focus on a functional operationalization of syntactic complexity allowed them to investigate different aspects of syntactic complexity in greater detail.

In terms of sampling conditions, it has been shown that syntactic complexity is influenced by a range of variables. While the educational institution and timed or un-timed writing conditions have been found to influence syntactic complexity measures in large scale corpus studies, the L1 has also been shown to explain variance in syntactic complexity of learners (Lu 2011: 50). Lu and Ai demonstrated in their study of English argumentative essays from learners of seven different L1 backgrounds that the same syntactic complexity measures yielded different results depending on the learners' L1 (2015: 25). It should be mentioned that this was also one of the few corpus studies on syntactic complexity that used L2 samples from learners who studied English as a foreign language (EFL) i.e. studying English in a non-English-speaking country. A specifically interesting finding from this study was that the German native speaker group achieved a higher score in syntactic complexity than the compared English native speaker group. This was particularly reflected in longer production units (i.e. sentences, T-units and clauses) of the German native speaker group compared to the ones found in the essays of the English native speaker group and the other L1 groups. Lu and Ai assume that a transfer from the L1 might have caused these differences between the two groups, because German is generally considered to be a language which tends to have longer sentences than English (25). The fact that the L1 as a modulator of syntactic complexity has only been investigated in a few L2 studies, implies that more research investigating the impact of this variable on syntactic complexity could lead to new insights regarding L2 writing (Ortega 2015: 85).

So far, the mentioned findings draw on data from the text genre of argumentative essays in an academic context. Interestingly, Yang Weiwei et al. found an influence of the writing topic on syntactic complexity measures in their study of argumentative essays which compared two different writing topics for L2 learners (2015: 65). While the topic requiring causal reasoning (i.e. the justification of why events follow each other) led to a higher use of subordination, the essay topic concerned with pure descriptive elements did not. Additionally, it has been shown that the field of academic writing in particular employs more syntactically complex language and relies on a higher use of nominalization and complex noun phrases than, for instance, narrative texts (Crossley et al. 2014: 204; Mazgutova & Kormos 2015: 7). Moreover, Beers and Nagy found differences in syntactic elements between the genres of persuasive essay and narrative texts of adolescent L1 writers (2009: 196, 2011: 197). They also mentioned that differences in syntactic complexity caused by writing genres depend on the construct operationalization. Nevertheless, an impact of the genre type on complexity was found between argumentative and narrative essays from college level L2 learners for 13 different complexity measures (Lu 2011: 48). The same study supports the argument that the genre of argumentative essay employs more syntactically complex structures than narrative texts. This demonstrates the importance of considering the text genre in studies on syntactic complexity. Ortega summarized these findings and proposed to consider not only different genres, but also previous genre experience of L2 writers for further studies of syntactic complexity (2015: 88, see also chapter 2.3).

In essence, this brief overview of studies points out various factors that influence syntactic complexity in L2 writing. At the same time, it illustrates the challenges involved in measuring syntactic complexity as one particular reflection of L2 development. The following methodological part of this thesis presents an attempt to measure syntactic complexity in the academic writing of L2 learners of English and aims to analyze some of the influencing factors.

5 Research questions

It has been pointed out in the theoretical part of this thesis that the use of complex language is essential in the process of putting thoughts into writing. Moreover, syntactic complexity is one aspect of language that is subject to change during L2 development and yields important information about the state of the L2 in the learner's mind. The investigation of syntactic development does not only contribute to the development of cognitive L2 writing models, but also leads to a better understanding of influencing variables for various applications in SLA research. While syntactic complexity in L2 academic writing has been widely studied in L2 learners studying in English speaking countries, there is little research about syntactic complexity in learners at universities outside of English speaking countries (see 4.4). Therefore, the overall aim of this thesis is to measure syntactic complexity in the academic papers of EFL students at the English department of the University of Vienna and to investigate if syntactic complexity might change over the course of their studies. For this reason, the first research question has been formulated:

(Q1) *Does syntactic complexity change in the academic writing of students at the English department of the University of Vienna between their first linguistic seminar paper and their last linguistic seminar paper?*

The hypothesis to this research question is that there is a change in syntactic complexity similar to previously reported developmental patterns of syntactic complexity for L2 writers from other institutions. It is expected that syntactic change is small in magnitude due to the high proficiency level and that the clausal and phrasal levels are the main areas of change. This assumption is based on the findings from previous research for learners with advanced English proficiency (Norris & Ortega 2009: 562–4; Ortega 2015: 89).

Furthermore, parts of the previous literature review pointed out that writing instruction is one extrinsic factor that influences syntactic complexity in the academic writing of L2 learners. This has been investigated primarily at lower and intermediate proficiency levels. It is, moreover, assumed that EAP courses trigger changes in syntactic development only at certain proficiency levels. Although short term academic writing courses showed minor influence on syntactic development at higher proficiency levels in past studies, this has only been investigated for periods of intensive instruction up to 4 months (Shaw & Liu 1998: 244).

Additionally, non-significant changes in measures specific to the genre of academic writing were reported by a similar study after four months of EAP instruction (Mazgutova & Kormos 2015: 12). Therefore, the second research question has been formulated to explore the effects of EAP courses on syntactic complexity over longer time spans:

(Q2) Does the number of academic writing courses taken between two linguistics courses account for differences in syntactic complexity?

It is hypothesized that the number of writing courses influences syntactic complexity in the investigated learner group. In contrast to previous studies, which did not support this hypothesis, the effect of EAP instruction is investigated over a longer time period than four months, and might therefore show an impact on syntactic development.

Besides writing instruction, other factors, such as the learner's L1, the text genre and the proficiency level have been shown to influence syntactic complexity in writing (Lu 2011). Although different genres have been investigated, the impact of previous genre experience on syntactic complexity has not yet been investigated in the L2. Nevertheless, it is assumed to influence the later syntactic development of L2 writers (Ortega 2015: 88). Genre experience in academic writing is usually first encountered towards the end of secondary education in Austria. In Austria, this depends to some extent on the type of secondary school. There are the two major upper secondary school types in Austria, namely the 'Allgemeinbildende Höhere Schule' (AHS) and the 'Berufsbildende Höhere Schule' (BHS). While the upper secondary part of the AHS lasts four years and mainly concentrates on preparing students to study at tertiary educational institutions, the BHS lasts five years and prepares students for different vocational fields according to the sub-type (BMB 2015, 2016). Particularly, the use of certain text genres in language classes is stated differently in their curricular description. The curricula of the AHS explicitly mentions the use of some academic writing genres, while the BHS curricula only explicitly states the teaching of genres connected to the vocational field (BKA 2017a, 2017d, 2017c, 2017b). Considering this and the assumed influence of previous genre experience on further syntactic development, the third research question has been formulated as follows:

(Q3) Do dimensions of syntactic complexity differ in the academic writing of EAP learners in depending on the type of secondary school they attended in Austria?

The hypothesis is that learners who went to an AHS show different outcomes in the analysis of syntactic measures than learners who went to a BHS, where less contact with academic writing in language education is assumed. Particularly, genre specific measures such as the MLC, the NFE/C or the CP/C are expected to be higher in the AHS group.

6 Materials and methods

6.1 Data collection and modification

The corpus for this study has been created from seminar papers of students who took two linguistics courses at the English department of the University of Vienna taught in English. Each data sample consists of extracts from two different academic papers and is accompanied by learner background information that has been elicited with a questionnaire. The first paper has been written at the beginning of their studies as part of an introductory course to linguistic research (i.e. Proseminar (PS) paper) and consists of around 2000 words. The second paper has been written as part of their last linguistics course which is usually taken towards the end of the degree program (i.e. Seminar (SE) paper) and comprises approximately 6000 words. Although learners have a deadline to hand in the papers, they are usually given enough time in between topic announcement and hand-in date. For this reason, both papers can be considered to have been written in an un-timed condition with enough time for planning and preparation. To collect the learner data, all participating students were given the questionnaire at the beginning of their last linguistics course. After collecting the filled out questionnaires, each student received a sheet with instructions to send their two papers to an e-mail address. The participation in the collection process was voluntary. In total, 107 questionnaires were handed out. In the end 61 full samples, each sample consisting of the learner's background information, one PS paper and one SE paper, have been collected. From these, 58 could be used for the analysis. A copy of the questionnaire and the instruction sheet can be found in the appendix (App. 1, pp. 81-3).

The received papers were manually converted from PDF (.pdf), Microsoft Word (.docx) or OpenOffice Writer (.odt) files into raw text files (.txt) for further processing with the text editor *gedit* (version 3.18.3). Headings, lists, tables and figures were manually removed during this step. Some papers included phonemic examples which have been replaced by the corresponding Latin letters. Furthermore, in-text citations that were longer than two lines were removed. In the next step, in-text references, all content written in parenthesis and quotation marks were automatically removed with the help of a script. The same script also replaced some special characters such as ampersands or the per cent sign with the corresponding words and inserted one space character and a new line after each sentence.

The applied script with comments (App. 3, p. 86) and sample paragraphs (App. 4, pp. 87-8) that illustrate the manipulations can be found in the appendix. However, the extracted texts have not been corrected or evaluated in any form for their accuracy. Furthermore, it was not possible to evaluate the time each participant spent on writing his or her paper; thus fluency could not be evaluated either. Consequently, the obtained text files then were cut to approximately 2000 words, at the end of the last sentence that was closest to this limit, resulting in the corpus specified in table 1.

Table 1: Descriptive statistics of the corpus

	Number of samples	**Mean [words]**	**SD [words]**	**Minimum [words]**	**Maximum [words]**	**Range [words]**	**Total amount of words**
PS papers	58	1962.5	60.1	1723	2011	288	113827
SE papers	58	1990.9	14.9	1934	2025	91	115473
							229300

Overall, table 1 displays that the corpus consists of 229300 words, with 58 PS papers ranging from 1723 words to 2011 words and 58 SE papers ranging from 1934 words to 2025 words. The parts extracted from the longer SE papers were taken from the first part of the paper, because these comprised the theoretical part and resemble the PS papers which are meant to treat a linguistic topic theoretically. Table 1 also shows that not all PS papers comprised 2000 words, as a result of the text modifications. This explains the higher range of the PS paper samples compared to the range of the SE paper samples.

6.2 Participants and sample information

The sample consisted of 58 students (48 female), aged between 20 and 47 years with a mean (M) of 24.4 years and a standard deviation (SD) of 3.9 years, as shown in table 2.

Table 2: Descriptive statistics of the learners

	Mean	**SD**	**Median**	**Min**	**Max**	**Range**
Age [years]	24.4	3.9	23	20	47	27
Formal English education [years]	9.4	1.9	9	5	13	8
Time at university [semesters]	8.6	2.1	9	4	14	10
Time between the two papers [semesters]	4.3	2.2	4	2	14	12

Writing courses between PS and SE paper	2.7	1.1	3	1	4	3

On average, students spent 8.6 (SD = 2.1) semesters at university until they started to write their SE paper, as shown in table 2. The average time interval between writing the PS paper and the SE paper was 4.3 semesters (SD = 2.2), and students took on average 2.7 (SD = 1.1) writing courses during this time. In terms of the L1, 48 learners spoke German as their L1, while 4 participants stated Bosnian, Croatian or Serbian (BCS) and two participants stated Ukrainian as their L1. The languages Polish, Italian, Spanish and Slovakian have each been stated by one learner as their parental language. Furthermore, the Polish and one BCS learner indicated German as their second L1 and therefore were considered as bilinguals. An overview of the L1s of the learners and other additional corpus information can be found in the appendix (App. 2, Table 9, pp. 84-5). Despite possible influences that might be caused by the different L1s, all participants speaking another L1 than German and the bilingual speakers have been kept in the sample to be representative for the population of interest at the English department of Vienna University. The influence of the L1, however, has been considered in some of the statistical analyses conducted and will be briefly discussed further in the results.

Previous secondary school types and academic writing education

Table 2 also displays that the learners had at least 5 years of formal English education prior to University, and the average learner had 9.4 years (SD = 1.9) of formal English. The participants, however, vary in type and amount of secondary education English. This is mainly caused through the diversity of Austria’s school system (see chapter 5). Thirty-six participants indicated that they went to an AHS and 19 participants reported a BHS as secondary school type to receive their high school leaving degree. Only 4 students went to a secondary school outside of Austria.

Considering the amount of language education of both school types, English and another foreign language are compulsory at the AHS, whereas in some BHS types English is the only foreign language being taught. Language education at the BHS, however, is again depending on the vocational field for which the BHS sub-type is specializes. For instance, a BHS for tourism or business includes compulsory language education in another foreign language besides compulsory English education, while those specialized in technical fields only teach English as compulsory foreign language. However, some BHS types – particularly those with the lowest compulsory English lessons per week – explicitly mention the use of Content and Language Integrated Learning (CLIL) in their curricula, but the BHS sub-type specific curricula vary in their explicitness about the amount of CLIL lessons per week. The different BHS sub-types have not been assessed in the questionnaire. Similarly, in the AHS the focus on language education can either be emphasized or shifted towards more lessons in the natural sciences or economics, but it is usually always higher or at least the same as the average amount of L2 education at a BHS. More importantly, the curriculum for the AHS explicitly states the teaching of complex language structures and their training. Furthermore, it states a wide range of text genre exposure, partly including academic writing genres in L2 education. The teaching of complex language is only implicitly mentioned in the different BHS curricula as part of general language proficiency specifications. Additionally, the text genres mentioned in the BHS curricula are connected to the respective vocational field (BKA 2017a, 2017b, 2017d, 2017c).

The curricula of all school types, however, state an English proficiency level of at least B2, according to the Common European Framework of Reference (CEFR), to receive the school leaving degree in Austria. More importantly, to enter the linguistics courses at the English department of Vienna University mentioned before, all of the participants had to pass the Common Final Test (CFT) at the end of their first two academic writing courses. The CFT tested the learners' English to ensure proficient reading and writing skills in academic contexts at a proficiency level of B2+ to C1 (CFT 2014: 1). All students are recommended to enter the PS course in the semester after they passed the CFT. Considering the continuing language courses and lectures in English after and during the PS course, a highly advanced proficiency level in academic writing can be assumed for most of the learners by the time they wrote their SE paper. As shown in table 2, the participants took at least one writing course between writing the two papers. Although each of the courses focuses on different aspects of academic writing, but not specifically on the area of syntax, no distinction was made between the type of courses. Furthermore, it is also assumed that the participants actively worked on the continuing improvement of their academic writing skills. Additionally, the learners usually have to take lecture exams or other degree specific courses in which they had at least some exposure to academic writing. It should be mentioned that learners taking both seminars usually express an affection towards English and language in general. Therefore, they could be deemed as more motivated to improve their English writing skills than learners who have to take a compulsory academic writing course as part of a non-language degree program.

6.3 Syntactic complexity measures

The extracts of the seminar papers were assessed for syntactic complexity with the measures used by Yang et al. (2015: 54) in their operalization of syntactic complexity (see chapter 4.3, p. 33). The only difference to their multi-dimensional construct is that the sub-construct of noun-phrase complexity was not applied in this study, due to problems in the utilization of this construct in the analysis. Table 3 presents a summary of the sub-constructs used, the syntactic level they tap and the applied measures and definitions.

Table 3: Definition of measures according to sub-constructs

Sub-construct	**Syntactic Level**	**Measure**	**Definition**
Overall sentence complexity	G	Mean length of sentence (MLS)	# words / # sentences
Overall T-unit complexity	G	Mean length of T-unit (MLTU)	# words / # T-units
Clausal coordination	S	T-unit per sentence (TU/S)	# T-units / # sentences
Clausal subordination (finite)	S	Dependent clause per T-Unit (DC/TU)	# dependent clauses / # T-units
Elaboration at clause level	C & P	Mean length of clause (MLC)	# words / # clauses
Phrasal coordination	C & P	Coordinate phrase per clause (CP/C)	# coordinate phrases / # clauses
Non-finite elements/subordination	C & P	Non-finite elements per clause (NFE/C)	# non-finite elements / # clauses

G = global, S = sentence, C = clause, P = phrase; ‘#’...number of, ‘/’...divided by

The software tool used for the analysis yielded the number of sentences, T-units, clauses, dependent clauses, coordinated phrases, verb phrases and words for each paper. From these sequencing units each of the syntactic complexity indices has been calculated as defined in table 3 with *LibreOffice Calc* (version 5.1.6.2). To arrive at the number of non-finite elements per clause, the number of verb phrases in a paper has been divided by the number of clauses and subtracted by 1. The reason for this is that each clause comprises one finite verb phrase according to the definition; thus subtracting 1 from the verb phrase per clause ratio yields the result for the remaining, non-finite verb phrases[7] (Yang et al. 2015: 59). All sequencing units were assessed in adherence to the definitions presented in chapter 4.3 with a software tool.

6.4 Analysis software

For the analysis of the sequencing units the *L2 syntactic complexity analyzer* (L2SCA) software was used. The L2SCA (version 3.3.3) is an open source software tool provided and described by Lu (2010, 2016). It was chosen because it is particularly designed for syntactic complexity assessment of advanced L2 writing. The whole software package has been downloaded from the homepage, including the *Stanford Parser* (version 3.3.1 see Klein & Manning 2003; Socher, Bauer, Manning, & others 2013) and the syntactic pattern match software *Tregex* (version 3.3.1 see Levy & Andrew 2006) that are implemented in the L2SCA (Lu 2016). The package has been run on *Ubuntu Linux* (16.04) in *batch mode* which allowed to analyze all files at once. For each paper extract the software first parses the plain text of the input raw text file and returns a sequence of parse trees. Each tree represents the syntactic structure of a sentence in the paper extract. An example of such a parse tree can be found in the appendix (App. 6, p. 100). The parse trees are then matched for the relevant syntactic structures by the software and the L2SCA returns the count of the sequencing units and the calculated indices for each input file (Lu 2010: 478).

The overall processing time of the 118 paper extracts with the L2SCA took around two hours without data preparation, while one manual analysis by a human rater of one paper took approximately six to eight hours. These numbers illustrate that the use of an analysis software provides great advantages such as a higher speed and more reliability. In comparison, human raters have to be trained for coding and there is the chance of subjective influence on ratings, despite clear coding guidelines (Crossley & McNamara 2014: 69). Nevertheless, some researchers still judge the use of automated coding tools as too inflexible for an accurate assessment of L2 production (Bulté & Housen 2014: 48).

Although these doubts are mainly concerned with the writings of lower proficiency levels or genres other than academic writing, the analysis software used in this paper has been compared to the human coding of three random samples from the analyzed corpus, which represents approximately five per cent of the whole corpus. They have then been compared for inter-annotator agreement by applying the suggestions of Brants (2000), in the same way as they have been used by Lu to validate the L2SCA (2010: 486).

---

[7] Subtracting the number of clauses, which equals the finite verbs in a paper, from the counted verb phrases and dividing by the number of clauses, yields mathematically the same value as dividing all verb phrases (finite and non-finite) by the number of clauses and subtracting 1.

Accordingly, the *precision score* has been computed with formula (5) shown below. It represents the ratio between the number of identical counts (i.e. the sequencing units found by both coding procedures) and the sequencing units found only by the human annotator. The *recall score* then is the ratio between identical sequencing units and the units found by the L2SCA, as shown in (6). Furthermore, the *F-score* represents the harmonic mean between recall and precision. It has been computed by using the formula in (7) and reflects the deviations of both coding procedures from each other. For this reason, it can be considered the most meaningful score to compare both procedures and judge their overall accuracy in detecting the desired units.

$$(5)\,Precision = \frac{number\ of\ identical\ structures\ found\ by\ the\ L2SCA\ \text{and}\ the\ annotator}{number\ of\ structures\ found\ by\ the\ annotator}$$

$$(6)\,Recall = \frac{number\ of\ identical\ structures\ found\ by\ the\ L2SCA\ \text{and}\ the\ annotator}{number\ of\ structures\ found\ by\ the\ L2\ SCA}$$

$$(7)\,\text{F-score} = \frac{2 * Precision * Recall}{(Precision + Recall)}$$

The identified sequencing units of both procedures and the calculated scores can be seen in table 4. Additionally, the texts (App. 5, pp. 88-99) and the rating comparisons of the two procedures for each sentence can be found in the appendix (App. 7, Tables 10 to 15, pp. 100-12, for a summary see App. 8, Table 16, p. 113).

Table 4: Sequencing unit identification between annotator *and L2SCA*

Identified	**Counts**			**Annotator-L2SCA agreement** [a]		
unit	**Annotator**	**L2SCA**	**Identical**	**Precision (Annotator)**	**Recall (L2SCA)**	**F-score**
Sentence	449	449	449	1	1	1
T-unit	526	504	502	.954	.996	.975
Clause	945	942	923	.977	.980	.978
Dependent Clause	419	413	392	.936	.942	.942
Coordinate Phrase	361	350	348	.964	.994	.979
Verb phrase	1268	1290	1255	.990	.973	.981
Words	11849	11849	11849	1	1	1

a) A score of ‘1’ means that there is no difference between identical sequencing units and the ones detected by the outlined procedure. The F-score describes the deviation of both procedures from the identical counts best, as it represents the harmonic mean of Precision and Recall.

As table 4 shows, sentences and words have been all coded identically by the human annotator and the L2SCA. The highest, but still marginal, difference between the two coding procedures can be seen for the T-units, with an F-score of .975. This is followed by higher deviations of the dependent clauses (DC), with an F-score of .942, and coordinate phrases (CP), with an F-score of .979. At the same time, these two sequencing units show the lowest identical counts in all samples (DC = 392 and CP = 348), which therefore puts a higher weight on their deviations.

In general, however, the F-score for every coded unit depicts that both procedures do not differ critically from each other. Lu used the same procedure to establish the validity and reliability of the L2SCA for L2 academic writing and the outcomes can therefore be compared (Lu 2010: 487–8). Except for the T-unit, the F-scores in Lu's findings are slightly lower, but show a similar pattern as the values in table 4. Although the number of different samples is higher in Lu's study (20 samples), the overall amount of analyzed sentences is smaller than in the present evaluation (323 compared to 449 sentences). Moreover, different text genres and probably different proficiency levels might have lead to further differences. Additionally, there have been two human annotators involved for the human annotations in Lu's evaluation, which increases the reliability.

The relatively small differences in the ratings in Lu's study are explained by parsing errors (e.g. annotating a verb phrase as a noun phrase) as major source of deviations, particularly involving the level of attachment (i.e. falsely parsed phrases lead to matching with wrong syntactic structures) and conjunctions (488). The same has been observed in the comparison conducted for this thesis. To trace possible errors, each sentence of the human annotations has been compared to the output of the Stanford parser (Klein & Manning 2003) that is part of the L2SCA. In general, it can be said that parsing errors of verbs (i.e. finite verbs parsed as non-finite verbs or other phrases) caused the most deviations between the two coding procedures, while some cohesive devices inhibited the correct parsing of clauses. Only few deviations could not be explained in this way. These discrepancies might have been due to mismatches by the L2SCA or misinterpretations by the human annotator. Deviations and possible reasons have been noted for each instance in the ratings and can be found in the appendix (App. 7, Tables 10 to 15, pp. 100-12).

As explained in the previous sub-chapter, the sequencing units discussed above have been used to calculate the actual measures of syntactic complexity. The sequencing units found by the annotator and the L2SCA each have been used to calculate the described complexity measured. This yielded two sets of measures; one set derived from the L2SCA and one set derived from the annotator sequencing units. The Pearson product-moment correlation has been calculated between both sets of measures and can be seen in table 5.

Table 5: Pearson's correlation coefficient (r) between annotator and L2SCA measures

Measure	MLS	MLTU	MLC	TU/S	DC/TU	CP/C	NFE/C
Pearson's *r*	1**	.988**	.995**	.912*	.993**	.997**	.839*

**significant at $p \leq .01$, *significant at $p \leq .05$

The two coding procedures significantly correlate for all measures, as table 5 shows. Pearson's *r* for each measure between both annotation methods ranges from .839 for the two NFE/C values to 1 for the MLS values. The correlations have been calculated in RStudio (version 1.0.143) and more detailed results can be found in the appendix (App. 8, Table 17, p. 113).

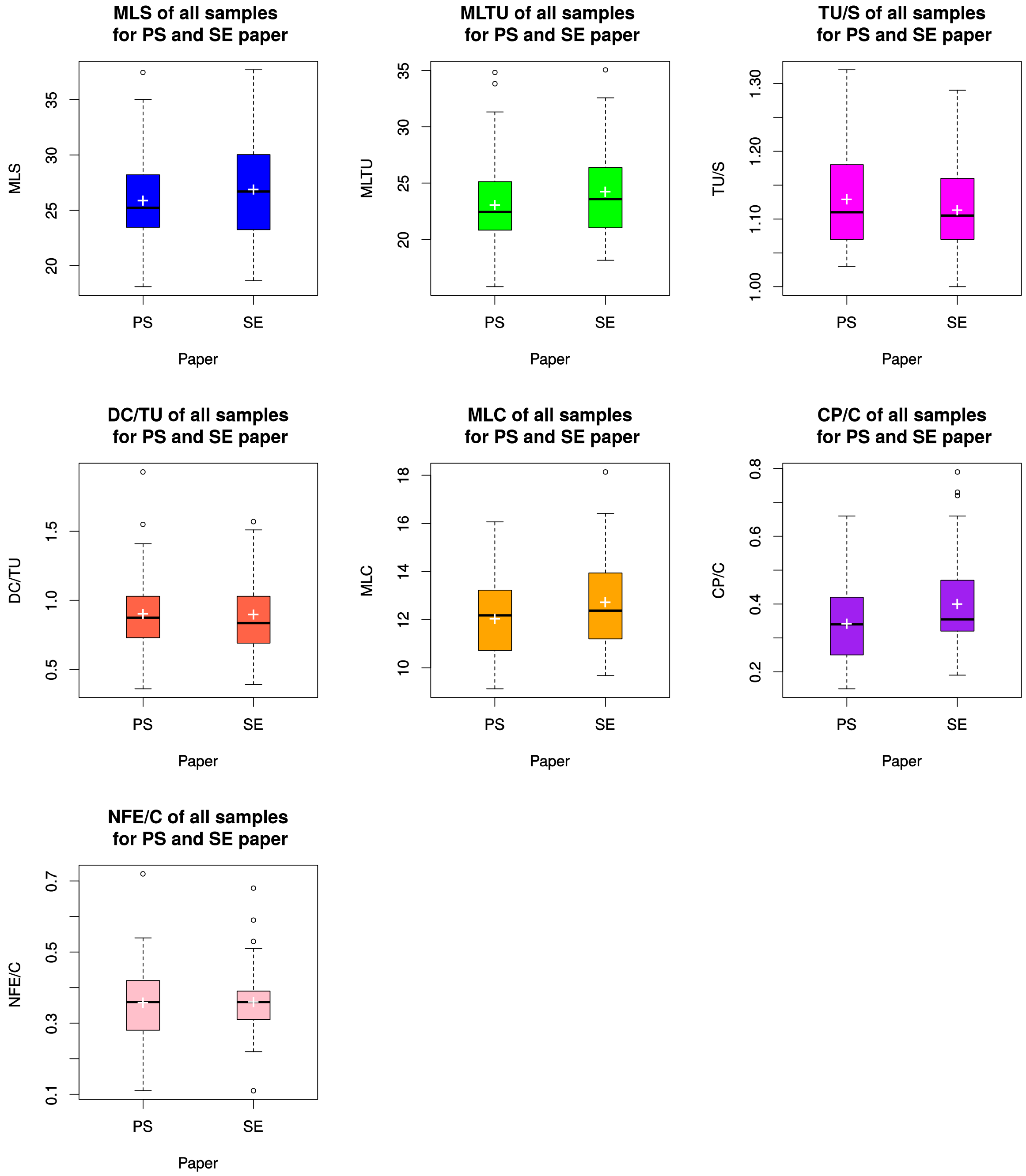
Overall, these results suggest that there is a negligible difference between the human rating and the L2SCA for the investigated data samples. Moreover, the comparison together with the enormous time that can be saved by using the L2SCA seems to justify the application of this software tool to calculate the defined measures of syntactic complexity. For this reason, the analysis of the whole corpus has been conducted with the L2SCA and the results for each sample and measure can be found in the appendix, including descriptive statistics for the whole data set (App. 9, Tables 18 to 20, pp. 114-9).

7 Results

The measurement results from the software analysis have then been used to answer the research questions with the help of statistical methods. All of the statistical tests carried out to answer have been calculated with RStudio (version 1.0.143) and the results for the test assumptions of the applied statistical tests can be found in the appendix (App. 10-11, pp. 119-121). A note on statistical significance: it has been argued that inferential statistical results in L2 research often lead to false interpretations in explaining observed relationships between variables or fail to even notice such relationships due to its dependence on sample size. For this reason, the effect size, expressed by Cohen's *d* or by η^2, has been calculated for the respective statistical tests results. This allows not only conclusions about observed changes that did not reach significance at the provided sample size, but also about the magnitude and the direction of an effect and therefore allow comparisons with other studies of different sample size. Notwithstanding the use of the effect sizes to indicate observable changes in the present sample, it should be emphasized that without clear indication of a significant result there is still the possibility that a reported effect might have been encountered by chance. As an estimation of this possibility the 95% confidence intervals (95%-CI) are always provided for Cohen's d (Cohen 1988; Norris & Ortega 2000: 493–5).

7.1 Q1: Change between the two seminar papers

The distribution of the measurement values obtained from the two papers for each of the seven measures are shown in the boxplots in figure 5. The type of paper is always plotted on the horizontal axes, while on the vertical axis of each plot the distribution for the respective measure can be seen. Furthermore, the bold vertical lines represent the median and the white plus-signs depict the arithmetic means of the plotted distribution accordingly. The upper and lower boundary of each of the boxes indicate the respective 75 % and 25 % quartiles. Moreover, the end of the whiskers – the dashed vertical lines extending from each side of the box – indicate the lowest and highest measured values that are not more than 1.5 times of the interquartile range (i.e. the range between the lowest and highest 25 % of samples in the distribution) away from the box in each direction. Consequently, the small circles outside of this range in some of the boxplots are single measurement values that represent statistical outliers. Outlying samples, however, have not been removed before the test, because of the high variation in measurement results of the outliers across all seven measures.


MLS of all samples for PS and SE paper

MLTU of all samples for PS and SE paper

TU/S of all samples for PS and SE paper

DC/TU of all samples for PS and SE paper

MLC of all samples for PS and SE paper

CP/C of all samples for PS and SE paper

NFE/C of all samples for PS and SE paper

Figure 5: Boxplots for all seven measures between PS paper and SE paper

It can be inferred from the different size of the boxes in figure 5 that the distribution of the measurement values for the 58 samples show some variation from the PS paper to the SE paper. For instance, the MLS (blue) and the MLTU (green) rose in variation, while the NFE/C (pink) clearly became more homogeneous between the two papers. Furthermore, the white plus signs in figure 5 indicate that the means of the global measures of syntactic complexity, represented by the MLS and the MLTU, show an overall rise from the PS paper to the SE paper for the 58 learners. This is also true for the specific and local level measures of the MLC (orange) and the CP/C (purple). Interestingly, the measure for sentential coordination, the TU/S (magenta), displays an overall decline in means from the PS paper to the SE paper. The two measures of sentential and clausal subordination, the DC/TU (red) and the NFE/C (pink), display no recognizable change.

To confirm these observations, two-tailed paired sample t-tests were applied to compare each of the seven measures of syntactic complexity between the PS paper and the SE paper. Each test compared the mean of the results obtained from the analysis of the 58 PS papers with the mean of the results from the corresponding 58 SE papers for the respective measurement value. The test results are shown in table 6. The first two columns in table 6 depict the investigated sub-constructs with their corresponding measures. In the next two columns the means of the 58 samples for each of the two papers are given with their standard deviations in parentheses for the respective measure. In the fourth column the computed *p*-values for the two-tailed paired sample t-tests between the PS and SE means with the corresponding t-value and the degrees of freedom in parentheses are given. Cohen's *d* as an indicator of the effect size with the 95%-CI for each value in parentheses are shown in the last column. Statistically significant results of the t-tests are marked with one asteriks.

Due to the fact that seven tests have been performed on the same data set, the Bonferroni correction[8] for the alpha value has been applied and therefore, the significance level for each of the tests has been set to $\alpha = .007$. Despite the controversial discussions about the Bonferroni correction, it seems justified to take on a more conservative significance level at the given sample size (Perneger 1998: 1237). Moreover, the Bonferroni correction has been applied in recent studies on syntactic complexity that used the same statistical methods (Lu 2011: 47; Lu & Ai 2015: 21).

---

[8] The Bonferroni correction is calculated by dividing the set alpha value by the number of tests performed on the same data set (e.g. set $\alpha = .05$, Bonferroni corrected $\alpha = .05/7 = .007$)

Table 6: Difference of syntactic complexity between PS paper and SE paper for all samples

Sub-construct	**Measure**	**PS mean (SD)**	**SE mean (SD)**	***p* (t-value/df)**	**Cohen's *d* ^{b} (95%-CI)**
Overall sentence complexity	MLS	25.79 (4.20)	26.79 (4.37)	.081 (-1.774/57)	.233 (.60/-.14)
Overall T-unit complexity	MLTU	22.96 (4.00)	24.14 (3.94)	.022 (-2.349/57)	.309 (.67/.06)
Clausal coordination	TU/S	1.13 (.073)	1.11 (.061)	.059 (1.923/57)	-.253 (.12/-.62)
Clausal subordination (finite)	DC/TU	.896 (.278)	.891 (.269)	.890 (.1385/57)	-.018 (.35/-.38)
Elaboration at clause level	MLC	12.00 (1.50)	12.68 (1.79)	.0006* (-3.621/57)	.476 (.85/.10)
Phrasal coordination	CP/C	.338 (.097)	.397 (.130)	.0011* (-3.438/57)	.451 (.82/.08)
Non-finite elements/subordination	NFE/C	.355 (.114)	.358 (.091)	.895 (-.1320/57)	.017 (.39/-.35)

*significant difference at $p \leq .007$
^{b} $d < 0$ indicates decline and $d > 0$ indicates rise in means from PS to SE paper for these values.

As can be seen in table 6, a rise in means of the MLC ($p = .0006$ / $t = -3.621$) and the CP/C ($p = .0011$ / $t = -3.438$) values from the PS paper to the SE paper is statistically significant. The significant change is also reflected in the effect sizes ($d_{MLC} = .476$ and $d_{CP/C} = .451$) and their 95%-CIs, which do not include zero. This means that on average, the number of words per clause significantly increased between writing the PS paper and the SE paper. At the same time, the amount of coordinated phrases used per clause is significantly higher in the SE papers than in the PS papers.

Despite the rise in means of the MLS and the MLTU from the PS paper ($M_{MLS} = 25.79$, $M_{MLTU} = 22.96$) to the SE paper ($M_{MLS} = 26.79$, $M_{MLTU} = 24.14$) shown in table 6, the t-test results for these measures suggest no significant difference. Consequently, this is also reflected in the smaller effect sizes for the MLS ($d_{MLS} = .233$) and the MLTU ($d_{MLTU} = .309$) compared to the strength of effects found for the MLC and the CP/C. However, the effect sizes still indicate an observable average rise between the PS paper and the SE paper. Although this difference is not significant at this sample size, the 95%-CIs for the d_{MLS} and the d_{MLTU} are clearly skewed in the positive direction, indicating a likely significant result at bigger sample sizes.

The DC/TU and the NFE/C clearly show no significant change in means between the two papers, but the means for clausal coordination (TU/S) indicates that the average coordination at sentence level applied by the learners seemed to decline from the PS paper to the SE paper. The t-test results do not yield statistically significant results for this difference in means for the TU/S and the effect is accordingly small with $d_{TU/S}$ = -.253. Nevertheless, the 95%-CI for $d_{TU/S}$ is here clearly skewed in the negative direction, indicating a higher likelihood to encounter a decline in coordination at sentence level for most samples.

In general, these results suggest that syntactic complexity changed between writing the PS paper and the SE paper in some dimensions with differing magnitude, while other dimensions showed no change at all. A point noteworthy to mention is the reported influence of the learner’s L1 in previous studies. All participants that had another L1 than German were kept in the sample to represent the average population of students at the English department of Vienna University. For reasons of consistency with other studies, however, the results for the same seven paired sample t-tests that were carried out only on the participants with German as their L1 can be found in the appendix (App. 13, Table 27, p. 122). Although these results do not differ in their significance from the results presented in table 6, their absolute means are higher and the magnitude of the observed effects differ, which could be caused due to the influence of the L1 German.

7.2 Q2: Influence of writing courses

The samples were then divided into groups of participants that had attended 1, 2, 3 or 4 writing courses between the PS and the SE. This was done to investigate the influence of the number of writing courses on the change in syntactic complexity. All measures that yielded significant results in the paired sample t-test conducted before have been investigated with the help of a one-way between groups analysis of variance (ANOVA). The number of courses figured as independent variable, while the difference between SE paper and PS paper was calculated for each sample and used as dependent variable for the ANOVA. The information of change between PS paper and SE paper is conserved in this value and referred to as Δ-value of the according measure. The significance level has been set to $\alpha = .025$ for each ANOVA, because two tests on the same data set have been carried out and the Bonferroni correction was applied. Table 7 shows the descriptive statistics for the Δ-values of the four groups.

Table 7: Group statistics of the one-way ANOVAs for Δ-MLS and Δ-CP/C

	No. of Courses (N)					
Statistics	**1 (9)**	**2 (18)**	**3 (17)**	**4 (14)**	**Total (58)**	**Measure[c]**
mean	1.05	.577	.706	.569	.683	Δ MLC
SD	2.07	1.46	1.31	1.16	1.44	
range	6.19	5.93	5.26	3.84	6.88	
median	.78	.581	.509	.537	.584	
mean	.071	.084	.066	.009	.058	Δ CP/C
SD	.135	.137	.131	.103	.127	
range	.376	.574	.560	.423	.719	
median	.021	.035	.071	-.011	.035	

[c] the Δ-value is the respective SE measure subtracted by the respective PS measure

The obtained groups display some variation in size, as only 9 learners took one course, while the groups with learners who took two or three courses are nearly double the size of group one. Furthermore, the high ranges in relation to the mean measurement differences and the respective SD for each of the four groups, indicate a high variation within each of the four groups for the MLC and the CP/C. This explains also the outliers that can be seen in the boxplots in figure 6, which show the Δ-value distributions for the two investigated measures.

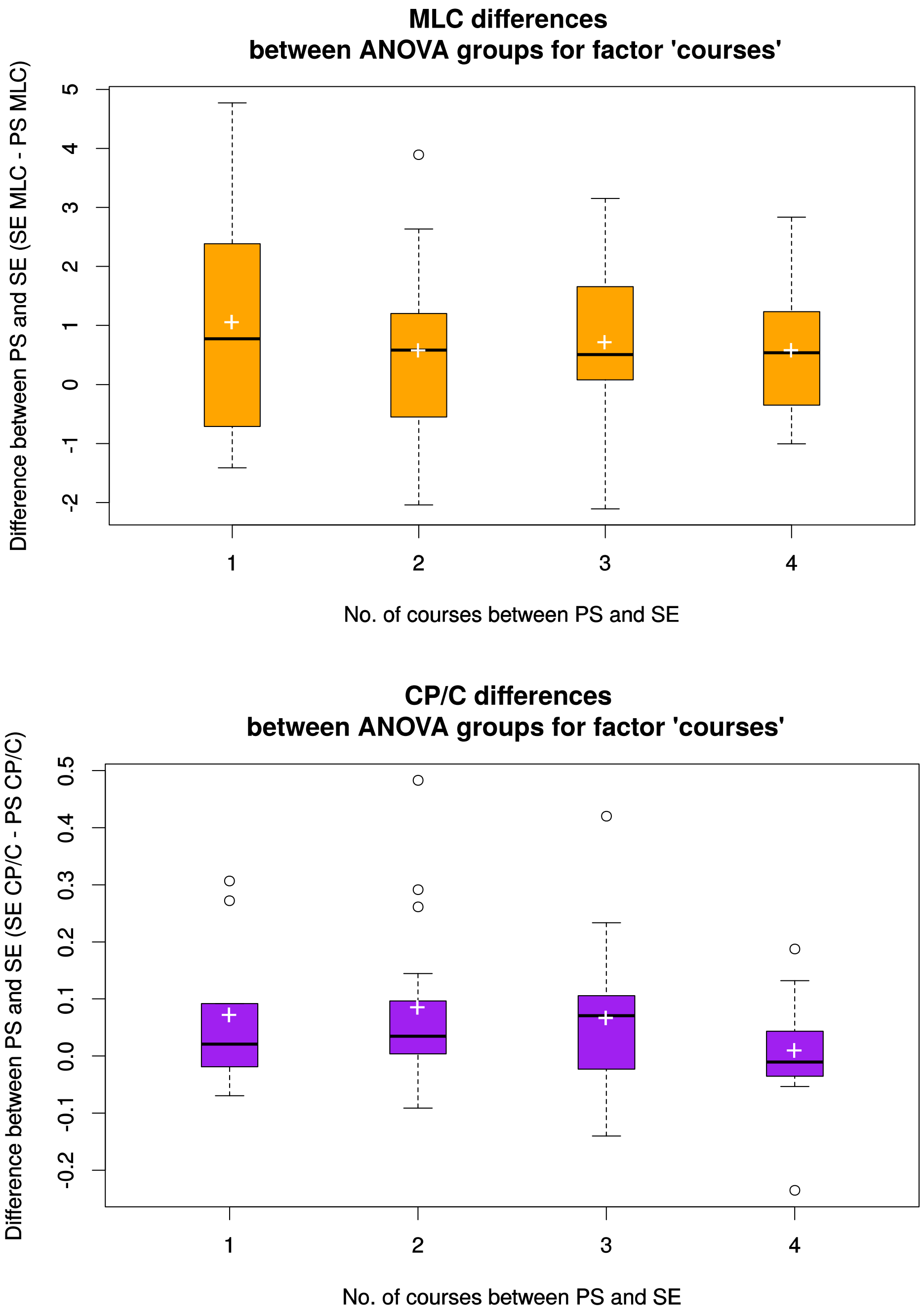


Figure 6: Boxplot for Δ-MLC and Δ -CP/C across number of writing courses

A removal of the outliers has not been conducted, because of the already small group sizes. Figure 6 also shows that the means do not suggest a general tendency for both measures across the four groups. The one-way between groups ANOVA in which the Δ-MLC values figured as dependent variable indicated no significant difference between the groups of learners who took 1, 2, 3 or 4 writing courses [F (3, 54) = .251, *p* = .860] with an effect size of $\eta^2 = .014$. Similarly, the ANOVA which investigated the Δ-CP/C values across the four groups with different amount of writing instruction found no significant difference between the groups [F (3, 54) = .397, *p* = .397] with an effect size of $\eta^2 = .053$. The very small effect sizes of both ANOVAs clearly suggest no tendency for change between any of the groups. The calculated between group and within group values for both ANOVAs can be found in the appendix (App. 14, Tables 28 & 29, p. 122).

Overall, these results suggest that the number of writing courses which were taken by the learners did not influence any of the measures that have significantly changed between writing the PS paper and the SE paper. Consequently, no further post-hoc tests were applied.

7.3 Q3: Influence of previous secondary education

The influence of previous secondary education was investigated with two-tailed independent sample t-tests to answer research question Q3. For this, the corpus has been split up into two groups of different size. The first group consists of 35 learners who attended the secondary school type of AHS and another group of 19 learners who went to a BHS. The seven tests carried out compared the means of the PS paper results only. This was done, because the PS papers were written closest to finishing secondary school, and therefore they were assumed to display less influence of writing instruction and genre experience at university level. At the same time, the influence of the L1 has not been considered in these tests. However, all of the learner samples used with a L1 other than German finished upper secondary school in Austria. The four learners who finished secondary school outside of Austria were excluded from the samples.

The boxplots in figure 7 present the distribution across the AHS and the BHS group for each of the seven measures. The first two boxplots show that the smaller BHS group displays a number of outliers for the overall measures of MLS (blue) and MLTU (green). This might be caused by the small sample size of the BHS group, because the boarders of the 1.5 times interquartile range of the larger AHS group show a similar spread in distribution for these measures. Moreover, the distribution of the values for the TU/S (magenta) and the MLC (orange) in the larger AHS group show a higher spread in distribution than in the BHS samples for these measures.

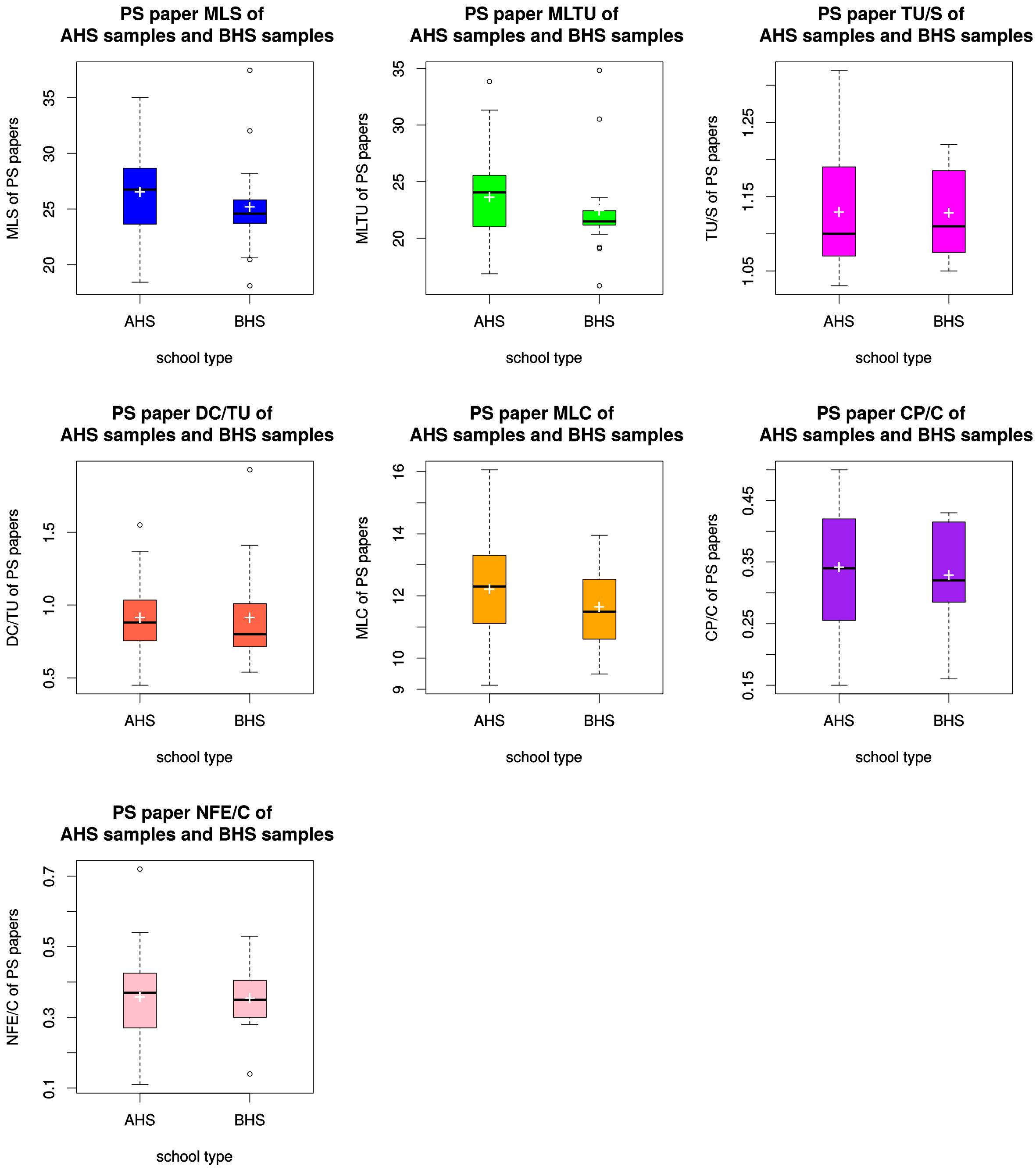


Figure 7: Boxplots for all seven PS paper measurement results between AHS and BHS

Nevertheless, the means of the MLS, MLTU, MLC and CP/C (purple) are slightly higher in the AHS group, while the means of all other measures are similar for both school types, as shown in figure 7.

These observable differences, however, did not reach the significance level for any of the measures in the independent sample t-test. The means and standard deviations with the according t-test results for each measure are presented in table 8. Furthermore, the significance level was set to $\alpha = .007$ for each test, because seven tests have been performed on the same data set.

Table 8: Differences of syntactic complexity between AHS PS papers and BHS PS papers

	AHS PS papers (N = 35)	**BHS PS papers (N = 19)**		
Measure	**mean (SD)**	**mean (SD)**	***p* (t-value/df)**	**Cohen's *d* [b] (95%-CI)**
MLS	26.44 (4.23)	25.10 (4.19)	.269 (1.117/52)	-.319 (-.89/.26)
MLTU	23.54 (4.02)	22.36 (4.08)	.313 (1.019/52)	-.289 (-.86/.29)
TU/S	1.128 (.084)	1.127 (.061)	.958 (.0531/52)	-.016 (-.59/.56)
DC/TU	.907 (.241)	.907 (.338)	.998 (-.0028/52)	.001 (-.57/57)
MLC	12.18 (1.58)	11.61 (1.29)	.184 (1.345/52)	-.407 (-.98/.17)
CP/C	.340 (.092)	.327 (.085)	.615 (.5062/52)	-.147 (-.72/.42)
NFE/C	.355 (.126)	.352 (.089)	.907 (.1179/52)	-.037 (-.61/.53)

[b] $d < 0$ indicates a decline and $d > 0$ indicates rise in means from the AHS to the BHS group for these values.

The results in table 8 show that the TU/S, DC/TU, CP/C and NFE/C indicate no tendency towards any difference between both groups. Although the general small sample size and the big difference in group size don not allow any reliable statements, the effect sizes for the MLS, MLTU and MLC suggest that the means differ for these measures between both groups. The effect size of the MLC ($d_{\mathrm{MLC}} = -.407$) might indicate that more syntactic elaboration at the clausal and phrasal level can be found in the AHS group. At least the 95%-CIs indicate a higher chance to encounter longer sentences, T-units and clauses in the AHS group compared to the BHS group for the investigated sample and group configuration. Nevertheless, these differences are still a matter of chance. A bigger sample size and more homogeneous group sizes are needed to form valid and reliable conclusions about the effects of the secondary school type on syntactic complexity.

8 Discussion

8.1 Q1: Change between two seminar papers

The present results suggest that performance in at least some dimensions of syntactic complexity change over time, more specifically between writing the two papers. Significant syntactic growth for this sample was observed at the clausal and sub-clausal level, as this is indicated by the rise in the two measures of MLC and CP/C. Regarding the MLC, on average the words in a clause rose significantly from the PS papers to the SE papers. This change can be partially explained by the CP/C, which shows that learners used significantly more coordinated phrases to raise the word count in a clause. The other sub-clausal measure that was applied was the NFE/C, measuring the number of non-finite constructions in a clause. Since this measure indicated no general tendency for change, the only explanation for the rise in clause length can be provided by coordination at the phrasal level. Of course, this does not mean that other sources of complexification at the sub-clausal level such as modification of noun phrases or more frequent use of adjuncts might not have contributed to this rise in the MLC. However, these other sources have only been tapped implicitly with the MLC.

The results of the other measures, despite being not significant at this sample size, strengthen the evidence that the main source of syntactic change between the two papers stems from the phrasal level. For instance, the global measures of MLS and MLTU indicate an overall tendency for a rise in words per unit of measurement. This growth at a global level cannot be traced back to the use of more subordination at sentence level, as the DC/TU remained unchanged. Moreover, the means of the measurements for coordination at sentence level, indicated by the TU/S, display a slight tendency for a decline between the two papers. This means that on average learners used less coordinated finite clauses, which should contribute to a decline of the MLS results. However, the opposite is the case and there are no other possible sources of complexification at the sentence level. Considering this and the significant results for the difference obtained for the MLC and the CP/C, it seems reasonable to conclude that the observed change in syntactic complexity solely happened at the clausal and phrasal level. Furthermore, it can only be said that on average non-finite constructions did not contribute to this growth, but phrasal coordination did significantly contribute to complexification at the sub-clausal level. These tendencies are in line with the proposed developmental pattern by Norris and Ortega, which predicted that syntactic growth mainly happens at the clausal and sub-clausal level for highly advanced learners. Moreover, their developmental pattern also predicts the non-evident change in both subordination measures, the DC/C and the NFE/C, while the observed decline in clausal coordination, expressed by the TU/S, is proposed from low proficiency levels onwards (Norris & Ortega 2009: 562). It should be noted that the investigated data set only delivers information about the predicted pattern towards the upper end of the proficiency scale and therefore, the results do not allow conclusions about syntactic development for low or intermediate proficiency levels.

Another observation that is specific to the investigated data set, concerns the absolute mean values of the global measure MLTU, which is seen as a valid indicator for any observed changes in syntactic development. In the investigated samples, the mean MLTUs of both papers are relatively high compared to the results reported in other studies. The same is true for most of the other measures. An explanation for this could be that global measures such as the MLTU generally tend to be higher in L1 German speakers compared to other L1 and native speaker groups in argumentative essays (Lu & Ai 2015: 22). This tendency is also indicated by the mean values of all measures for the L1 German samples listed in the appendix (App. 13, Table 27, p. 122). They are slightly higher than the mean values, which include the samples of speakers with other parental languages. Although this might provide support for the argument that L1 transfer of syntactic structures is very likely, the other L1 groups in the data set were too small to allow valid conclusions about this phenomenon (25). Considering the influence of the L1, comparable L1 German speaker groups with a similar proficiency level did still show much lower absolute mean values for the MLTU results in argumentative essays of other studies (Lu & Ai 2015: 22; Mazgutova & Kormos 2015: 11). Although it has been shown that syntactic complexity might vary across educational institutions (Lu 2011: 50), these relatively high differences might also be due to the specific sub-genre of academic paper that might employ even more complex language than the argumentative essays analyzed in most other studies. This comparison of the MLTU should illustrate the challenge to consider all the possible variables that might influence syntactic complexity.

It should also be noted that the outlined changes only display an overall tendency in the investigated data set for certain dimensions of syntactic complexity. Without questioning the results of the statistical analysis, it should still be mentioned that the individual samples often showed a different developmental pattern than the outlined general tendency suggests[9]. Although some samples exhibit lower values for some of the tapped constructs in the SE papers, it should always be considered that these results show only one, very narrow picture of a few syntactic facets taken at a specific point in time. Moreover, the theoretical part explained that complex writing is always subject to a range of other, often very individual factors which could not all be controlled in this corpus study. This should be considered particularly in the light of the theoretical underpinnings provided by the DST approach to investigate developmental L2 patterns on an individual basis. The samples not complying with the overall tendencies outlined can be taken as evidence that L2 development happens in individual and variable ways that are difficult to predict over time. More importantly, the changes in global syntactic complexity of individual samples can definitely not be taken as an absolute indicator for overall patterns of L2 writing development. Other dimensions of L2 development such as fluency, accuracy or lexical complexity might still change over time, while syntactic development declines or stagnates. It should also be noted that in some cases a decline in certain areas still indicates development, as in the case of the outlined pattern where the decline in coordination at sentence level is argued to indicate development. Moreover, this illustrates that it is essential to consider all three dimensions of CAF for investigating L2 development. Particularly, when higher emphasis is put on individual learner development such as in DST driven approaches. Consequently, this might offer better explanations for the discrepancies between the measures of some individual samples. However, to investigate syntactic complexity on a more individual basis, a different research design would have to be applied.

8.2 Q2: Influence of writing courses

Interestingly, the results of the ANOVA between the learner groups arranged according to the number of EAP courses showed no difference for the two tested measures of MLC and CP/C. Clearly, the results should by no means lead to the conclusion that EAP instruction does not enhance L2 development at higher proficiency levels. It can only be said that the number of courses taken between writing the two papers did not explain any significant difference in the development of overall sub-clausal complexification and coordination at phrasal level.

---

[9] The measurement values for sample 28 or sample 8 can be taken as examples for such a behavior. Their values for both papers and the difference between those are listed in the appendix.

Moreover, the applied construct does not allow any conclusions about other areas of L2 development that might have been influenced by the number of writing courses. These results support the outcome of other studies which came to similar results about advanced proficiency groups, but which investigated smaller time spans of instruction (Mazgutova & Kormos 2015: 11–12). Mazgutova and Kormos state that syntactic development at high proficiency levels exhibits very little change under the influence of EAP instruction that does not explicitly concentrate on the area of syntax. As already mentioned, Shaw and Liu came to similar conclusions about the influence of EAP instruction at higher proficiency levels (1998: 246). Their outcomes are clearly supported by the results obtained from the two ANOVAs in this thesis. Moreover, it can be said that even after a period of at least eight months, the assumed minimum in this study, no influence of EAP instruction on the tested sub-constructs can be seen.

Nevertheless, Mazgutova and Kormos also reported an increase in the use of complex nominals and noun phrase modifiers after EAP instruction (2015: 12). Although these constructs have not been measured in the present study, the results can be argued to indicate such an increase in phrasal elaboration as expressed by the significant rise in the MLC and the CP/C (see 8.1). However, these changes do not result from EAP instruction in the present study, according to the ANOVA outcomes. More specific sub-constructs of syntactic complexity might still indicate a difference that is caused by the number of writing courses.

Finally, it should also be mentioned that only the number of writing courses taken in the time between writing the PS paper and the SE paper were analyzed. I has not been considered at which point in time and which type of writing course was taken. These factors together with the inhomogeneous and small groups sizes might also have influenced the results.

8.3 Q3: Influence of previous secondary education

There was no significant difference in measures of syntactic complexity between students who had previously attended AHS and those who had previously attended BHS. Despite the non-significant results, the tendency observed for the global measures of syntactic complexity, the MLS and the MLTU, suggest that learners who attended a BHS did use less words per sentence or T-unit in their PS papers than their colleagues who went to an AHS. Interestingly, the effect size indicated that overall clausal elaboration (MLC) was also higher for previous the AHS students than for the previous BHS students. Although the CP/C and the NFE/C did not show any differences between the two groups, further investigation with bigger sample sizes might confirm the assumed difference in the MLC. Concerning the other coordination and subordination measures at sentence and phrasal level, there was no difference in means between the AHS and the BHS learners. Consequently, the overall trend derived and based on the reported effect sizes suggests that the lower amount of syntactic complexity applied by the BHS group might stem from the phrasal and clausal level.

From this, it could be tentatively argued that the difference between both learner groups might, at least partially, originate from complexification at the phrasal level. One could speculate that the difference in academic specific genre exposure in both school types might have influenced this difference that points towards genre specific sub-constructs. Nevertheless, conclusions drawn about genre exposure solely based on the secondary school type have to be considered very carefully, because other factors could have had a similar impact. For instance, the varying amount of L2 instruction in both school types or the influence of academic writing instruction at the university are variables that might contribute to the observed difference.

Above all, the sample size and the difference in group sizes are a problematic issue and some individual samples in the BHS groups show the same or even higher values as the ones in the AHS group. This could lead to very different outcomes at bigger sample sizes, particularly when considering that the different sub-types of BHSs have not been controlled in the BHS group. Moreover, it should be considered that certain individual factors could be applied to the BHS group. It could be assumed that learners in the BHS group usually prefer to study a subject connected to their vocational field or start working in this field over studying a language. Due to the learner’s choice to study English at university, a higher motivation to learn English or a general personal affection towards languages might be evident in all of the BHS group samples. Although this applies in a similar way to the AHS group, it might distort the results for the BHS group in a particular way.

Considering the outlined variables that have not been considered in the statistical analysis, the results do not allow for any particular and valid conclusions about the influence of the secondary school type on the investigated areas of syntactic complexity. Notwithstanding the insignificance of the results, the observed tendencies in the effect sizes of the global measures might still lead to interesting results at reasonable sample and group sizes in further research.

9 Conclusion & implications for further research

Regarding the first research question, it can be concluded that the results indicate that overall syntactic development occurs mainly at the phrasal and clausal level for the investigated learner group. This can be taken as support for syntactic developmental patterns predicted by previous research for advanced proficiency levels. The results, moreover, provide evidence that syntactic development in the investigated data set mainly stems from phrasal coordination, while other sub-clausal and phrasal dimensions that have not been analyzed might as well contribute to this change. This can be taken as a point of departure for further research in this field. On the basis of these findings, it is suggested for further research to concentrate on other modifications and sub-constructs at the phrasal and clausal level. Although change was found for some dimensions at this proficiency level, it seems logical to apply measures that tap other forms of phrasal complexification. In particular, measures of noun phrase complexity could provide more information about how complexity might be achieved in the investigated genre and at this proficiency level. Furthermore, indications for a possible L1 transfer of syntactic structures to L2 academic writing were found, but could not be validated due to a lack of necessary data. Overall, the results provide valuable information that can serve to investigate syntactic complexity in future learner groups or writing genres not only at the University of Vienna, but also across other educational institutions.

In contrast to what was hypothesized for the second research question, the number of EAP courses and the assumed longer period of instruction did not contribute to the significant changes found in phrasal coordination or overall clausal and sub-clausal elaboration. The influence of the number of EAP courses on other areas that are very likely to be influenced by EAP instruction such as lexical complexity, fluency or accuracy have not been investigated. Therefore it is proposed to consider these areas together in further research to investigate the influence of EAP instruction on L2 writing development more globally. Notwithstanding the limited scope of the investigated constructs, the results provided are in line with previous study outcomes that explored syntactic development at highly advanced proficiency levels over shorter time periods. The results suggest that syntactic development at the phrasal or clausal level does not seem to be influenced by EAP instruction. However, a bigger sample size and more homogeneous groups are needed to draw reliable conclusions and rule out any influence of EAP instruction on long-term syntactic development.

With regard to the third research question, it can be stated that the analysis did not allow valid conclusions about differences in syntactic complexity between school types. Nevertheless, the analysis showed that for some measures, there was a difference between learners who went to an AHS and learners who went to a BHS. Furthermore, it has also been discussed that various other factors such as the BHS sub-type or previous tertiary education might play a role here and should be considered in further research. As previously discussed, the results could indicate differences that are related to previous genre exposure, but further research is needed to investigate this variable. Additional research of previous genre exposure and language education might prove useful for the pedagogical implication and help teachers to raise awareness of such influences.

Despite the limited and specific scope of the present study, the results provide useful information for further research in the field of syntactic complexity. The analyzed data set might not only serve as benchmark data to investigate syntactic complexity in future student populations at the University of Vienna, but also serve as a starting point to explore overall syntactic development patterns at lower proficiency levels. Similarly, the results can be used to compare and investigate possible institutional differences among learners. Considering the research methodology of this thesis, the used analysis software might also find use to obtain a quick overview of research data in terms of genre specific indicators or approximate L2 development. The evaluation of the L2SCA might also help to improve linguistic analysis software and contributes to the validity and reliability of the LSCA. The L2SCA – and possible future software adaptations of it – might be useful not only in research, but also for teachers to track learner development or estimate improvements in some language areas.

A final word on the investigated data set and the limitations of the analysis. Although the results indicated overall developmental patterns of L2 learners' syntactic complexity, the thesis also demonstrated that the diverse learner backgrounds cannot solely be reduced to one simple trajectory describing L2 development in the area of syntax. This point is essential when interpreting the results of this study. Moreover, the results only depict a very specific and narrow dimension of L2 development in writing. Still, they serve as one piece in the puzzle to be solved for enlightening our understanding of L2 writing development.

References

Alamargot, Denis; Dansac, Christopher; Chesnet, David; Fayol, Michel. 2007. "Parallel processing before and after pauses: a combined analysis of graphomotor and eye movements during procedural text production". In Torrance, Mark; Waes, Luuk van; Galbraith, David (Eds.), *Writing and cognition: research and applications*. 13–29. Amsterdam: Elsevier.

Bachman, Lyle F. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Barnett, Bruce. 2015. "Sed - An introduction and tutorial". Retrieved 27 May 2017, from http://www.grymoire.com/Unix/Sed.html#uh-0

Baten, Kristof; Håkansson, Gisela. 2015. "The development of subordinate clauses in German and Swedish as L2s: A theoretical and methodological comparison". *Studies in Second Language Acquisition*, *37*(3), 517–547. http://doi.org/10.1017/S0272263114000552

Beckner, Clay; Bybee, Joan; Christiansen, Morten H.; Croft, William; Ellis, Nick C.; Holland, John; et al. 2009. "Language is a complex adaptive system: position paper". *Language Learning*, *59*, 1–26. http://doi.org/10.1111/j.1467-9922.2009.00533.x

Beers, Scott F.; Nagy, William E. 2009. "Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre?". *Reading and Writing*, *22*(2), 185–200. http://doi.org/http://doi.org/10.1007/s11145-007-9107-5

Beers, Scott F.; Nagy, William E. 2011. "Writing development in four genres from grades three to seven: syntactic complexity and genre differentiation". *Reading and Writing*, *24*(2), 183–202. http://doi.org/http://dx.doi.org/10.1007/s11145-010-9264-9

Berman, Ruth A.; Nir-sagiv, Bracha. 2007. "Comparing narrative and expository text construction across adolescence: A developmental paradox". *Discourse Processes*, *43*(2), 79–120. http://doi.org/10.1080/01638530709336894

Biber, Douglas; Gray, Bethany; Staples, Shelley. 2016. "Predicting patterns of grammatical complexity across language exam task types and proficiency levels". *Applied Linguistics*, *37*(5), 639–668. http://doi.org/10.1093/applin/amu059

BKA, Bundeskanzleramt. 2017a. "Gesamte Rechtsvorschrift für Lehrpläne – allgemeinbildende höhere Schulen". Retrieved 27 August 2017, from https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10008568&FassungVom=2017-08-31

BKA, Bundeskanzleramt. 2017b. "Gesamte Rechtsvorschrift für Lehrpläne - Handelsakademie und Handelsschule". Retrieved 27 August 2017, from https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10008944

BKA, Bundeskanzleramt. 2017c. "Gesamte Rechtsvorschrift für Lehrpläne - Höhere Lehranstalt für wirtschaftliche Berufe". Retrieved 27 August 2017, from https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10008878

BKA, Bundeskanzleramt. 2017d. "Gesamte Rechtsvorschrift für Lehrpläne der Höheren technischen und gewerblichen Lehranstalten". Retrieved 27 August 2017, from https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20007451

Bley-Vroman, Robert. 1983. "The Comparative Fallacy in interlanguage studies: The case of systematicity". *Language Learning*, *33*(1), 1–17. http://doi.org/10.1111/j.1467-1770.1983.tb00983.x

BMB. 2015. "Bundesministerium für Bildung - Allgemein bildende höhere Schule (AHS)". Retrieved 27 September 2017, from https://www.bmb.gv.at/schulen/bw/abs/ahs.html

BMB. 2016. "Bundesministerium für Bildung - Berufsbildende mittlere und höhere Schulen - Allgemeines". Retrieved 27 September 2017, from https://www.bmb.gv.at/schulen/bw/bbs/bmhs_allg.html

Bonin, Patrick; Fayol, Michel; Gombert, Jean-Emile. 1998. "An experimental study of lexical access in the writing and naming of isolated words". *International Journal of Psychology*, *33*(4), 269–286. http://doi.org/10.1080/002075998400312

Brants, Thorsten. 2000. "Inter-annotator agreement for a german newspaper corpus". In *LREC*. Athens, Greece. Retrieved from https://pdfs.semanticscholar.org/d630/9b4cbaf24f65dd9582d21397e5661567fcab.pdf

Bulté, Bram; Housen, Alex. 2012. "Defining and operationalising L2 complexity". In Housen, Alex; Vedder, Ineke; Kuiken, Folkert (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. 21–46. Amsterdam/Philadelphia: John Benjamins.

Bulté, Bram; Housen, Alex. 2014. "Conceptualizing and measuring short-term changes in L2 writing complexity". *Journal of Second Language Writing*, *26*, 42–65. http://doi.org/10.1016/j.jslw.2014.09.005

Cameron, Lynne; Larsen-Freeman, Diane. 2007. "Complex systems and applied linguistics". *International Journal of Applied Linguistics*, *17*(2), 226–240. http://doi.org/10.1111/j.1473-4192.2007.00148.x

Campbell, Cherry. 1990. "Writing with others' words: using background reading text in academic compositions". In Kroll, Barbara (Ed.), *Second language writing: research insights for the classroom*. 211–230. Cambridge: Cambridge University Press.

Caspi, Tal. 2010. *A dynamic perspective on second language development* (PhD thesis). University of Groningen. Retrieved from http://www.rug.nl/research/portal/publications/pub(cf0df3e6-9bc8-4b99-93b4-9dba694c9896).html

CFT, Department of English and American studies, University of Vienna. 2014. "CFT New: Test Specifications". Retrieved from http://archivanglistik.univie.ac.at/fileadmin/user_upload/dep_anglist/StudienServiceStelle/Organisatorisches/CFT_Online_info_Test_Specifications_March_2014.pdf

Chenoweth, N. Ann; Hayes, John R. 2001. "Fluency in writing: generating text in L1 and L2". *Written Communication*, *18*(1), 80–98. http://doi.org/10.1177/0741088301018001004

Chenoweth, N. Ann; Hayes, John R. 2003. "The inner voice in writing". *Written Communication*, *20*(1), 99–118. http://doi.org/10.1177/0741088303253572

Cleland, Alexandra A.; Pickering, Martin J. 2006. "Do writing and speaking employ the same syntactic representations?". *Journal of Memory and Language*, *54*(2), 185–198. http://doi.org/10.1016/j.jml.2005.10.003

Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences* 2. ed.. Hillsdale, NJ: Erlbaum.

Crossley, Scott A.; McNamara, Danielle S. 2014. "Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners". *Journal of Second Language Writing*, *26*, 66–79. http://doi.org/10.1016/j.jslw.2014.09.006

Crossley, Scott A.; Roscoe, Rod; McNamara, Danielle S. 2014. "What is successful writing? An investigation into the multiple ways writers can write successful essays". *Written Communication*, *31*(2), 184–214.

De Keyser, Robert M. 2009. "Cognitive-psychological processes in second language learning". In Long, Michael H.; Doughty, Catherine J. (Eds.), *The handbook of language teaching*. 119–138. Chichester, UK: Wiley-Blackwell.

Dörnyei, Zoltán. 2007. *Research methods in applied linguistics: quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.

Dörnyei, Zoltán. 2009. *The psychology of second language acquisition*. Oxford: Oxford University Press.

Ellis, Rod; Barkhuizen, Gary P. 2005. *Analysing learner language*. Oxford: Oxford University Press.

Gathercole, Susan E.; Baddeley, Alan D. 1993. *Working memory and language*. Hove, East Sussex: Erlbaum.

Givón, Talmy. 2009. *The genesis of syntactic complexity: diachrony, ontogeny, neuro-cognition, evolution*. Amsterdam/Philadelphia: John Benjamins.

Goyvaerts, Jan. 2016. "Regex Tutorial". Retrieved 1 March 2017, from http://www.regular-expressions.info/tutorial.html

Grabowski, Joachim. 1996. "Writing and speaking: common grounds and differences toward a regulation theory of written language production". In Levy, Micheal C.; Ransdell, Sarah (Eds.), *The science of writing: theories, methods, individual differences, and applications*. 73–91. Mahwah, New Jersey: Lawrence Erlbaum.

Halliday, Michael Alexander Kirkwood; Martin, James Robert. 1993. *Writing Science : Literacy and Discursive Power*. London: Routledge.

Hayes, John R. 2009. "From idea to text". In Beard, Roger; Myhill, Debra; Riley, Jeni (Eds.), *The Sage handbook of writing development*. 65–79. LA, California: Sage.

Hayes, John R. 2012. "Modeling and remodeling writing". *Written Communication*, *29*(3), 369–388.

Hayes, John R.; Olinghouse, Natalie G. 2015. "Can cognitive writing models inform the design of the Common Core State Standards?". *The Elementary School Journal*, *115*(4), 480–497.

Housen, Alex; Kuiken, Folkert. 2009. "Complexity, accuracy, and fluency in second language acquisition". *Applied Linguistics*, *30*(4), 461–473. http://doi.org/10.1093/applin/amp048

Housen, Alex; Kuiken, Folkert; Vedder, Ineke. 2012. "Complexity, accuracy and fluency: Definitions, measurement and research". In Housen, Alex; Vedder, Ineke; Kuiken, Folkert (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. 1–20. Amsterdam/Philadelphia: John Benjamins.

Hunt, Kellogg W. 1965. "Grammatical structures written at three grade levels". *NCTE Research Report No. 3*. Retrieved from http://eric.ed.gov/?id=ED113735

Hunt, Kellogg W. 1970. "Do sentences in second language grow like those in the first?". *TESOL Quarterly*, *4*(3), 195–202.

Kellogg, Ronald T. 1994. *The psychology of writing*. New York: Oxford University Press.

Kellogg, Ronald T. 1996. "A model of working memory in writing". In Levy, Micheal C.; Ransdell, Sarah (Eds.), *The science of writing: theories, methods, individual differences, and applications*. 57–71. Mahwah, New Jersey: Lawrence Erlbaum.

Klein, Dan; Manning, Christopher D. 2003. "Fast exact inference with a factored model for natural language parsing". In *Advances in neural information processing systems*. 3–10. Retrieved from http://papers.nips.cc/paper/2325-fast-exact-inference-with-a-factored-model-for-natural-language-parsing.pdf

Kormos, Judit. 2011. "Speech production and the Cognition Hypothesis". In Robinson, Peter (Ed.), *Second language task complexity : Researching the Cognition Hypothesis of language learning and performance*. 39–60. Amsterdam/Philadelphia: John Benjamins.

Kormos, Judit. 2014. "Differences across modalities of performance: an investigation of linguistic and discourse complexity in narrative tasks". In Manchón, Rosa; Byrnes, Heidi (Eds.), *Task-based language learning: insights from and for L2 writing*. 194–216. Amsterdam/Philadelphia: John Benjamins.

Kuiken, Folkert; Vedder, Ineke. 2011. "Task complexity and linguistic performance in L2 writing and speaking: the effect of mode". In Robinson, Peter (Ed.), *Second language* *task complexity : Researching the Cognition Hypothesis of language learning and performance*. 91–104. Amsterdam/Philadelphia: John Benjamins.

Larsen-Freeman, Diane. 1978. "An ESL index of development". *TESOL Quarterly*, *12*(4), 439–448.

Larsen-Freeman, Diane. 1983. "Assessing global second language proficiency". In Seliger, Herbert W.; Long, Michael H. (Eds.), *Classroom oriented research in second language acquisition*. 287–304. Rowley, Massachusetts: Newsbury House.

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

Larsen-Freeman, Diane. 2006. "The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English". *Applied Linguistics*, *27*(4), 590–619. http://doi.org/10.1093/applin/aml029

Larsen-Freeman, Diane. 2009. "Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition". *Applied Linguistics*, *30*(4), 579–589. http://doi.org/10.1093/applin/amp043

Larsen-Freeman, Diane. 2015. "Complexity Theory". In Van Patten, Bill; Williams, Jessica (Eds.), *Theories in second language acquisition: An introduction*. (2nd edition).227–244. New York: Routledge.

Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, Massachusetts: MIT Press.

Levy, Roger; Andrew, Galen. 2006. "Tregex and Tsurgeon: tools for querying and manipulating tree data structures". In *Proceedings of the fifth international conference on Language Resources and Evaluation*. 2231–2234. Retrieved from http://idiom.ucsd.edu/~rlevy/papers/levy_andrew_2006.pdf

Lowie, Wander; Caspi, Tal; Van Geert, Paul; Steenbeek, Henderien. 2011. "Modeling development and change". In Verspoor, Marjolijn H.; De Bot, Kees; Lowie, Wander (Eds.), *A dynamic approach to second language development: methods and techniques*. 99–121. Amsterdam/Philadelphia: John Benjamins.

Lowie, Wander; Verspoor, Marjolijn H. 2015. "Variability and variation in second language acquisition orders: A dynamic reevaluation". *Language Learning*, *65*(1), 63–88. http://doi.org/10.1111/lang.12093

Lu, Xiaofei. 2010. "Automatic analysis of syntactic complexity in second language writing". *International Journal of Corpus Linguistics*, *15*(4), 474–496. http://doi.org/10.1075/ijcl.15.4.02lu

Lu, Xiaofei. 2011. "A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development". *Tesol Quarterly*, *45*(1), 36–62.

Lu, Xiaofei. 2016. "L2 Syntactic Complexity Analyzer" [Software Distribution]. Retrieved 23 May 2017, from http://www.personal.psu.edu/xxl13/downloads/l2sca.html

Lu, Xiaofei; Ai, Haiyang. 2015. "Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds". *Journal of Second Language Writing*, *29*, 16–27. http://doi.org/10.1016/j.jslw.2015.06.003

Manchón, Rosa M.; Roca De Larios, Julio. 2007. "On the temporal nature of planning in L1 and L2 composing". *Language Learning*, *57*(4), 549–593. http://doi.org/10.1111/j.1467-9922.2007.00428.x

Manchón, Rosa M.; Roca De Larios, Julio; Murphy, Liz. 2009. "The temporal dimension and problem-solving nature of foreign language composing processes. Implications for theory". In Manchón, Rosa (Ed.), *Writing in foreign language contexts: learning, teaching, and research*. 102–129. Bristol: Multilingual Matters.

Mazgutova, Diana; Kormos, Judit. 2015. "Syntactic and lexical development in an intensive English for Academic Purposes programme". *Journal of Second Language Writing*, *29*, 3–15. http://doi.org/10.1016/j.jslw.2015.06.004

McCutchen, Deborah. 1996. "A capacity theory of writing: Working memory in composition". *Educational Psychology Review*, *8*(3), 299–325.

Mesibov, Bob. 2014, April 23. "Scripting a ‘find-and-replace’ for big text files". Retrieved 22 May 2017, from http://www.thelinuxrain.com/articles/scripting-a-find-and-replace-for-big-text-files

Miestamo, Matti. 2008. "Grammatical complexity in a cross-linguistic perspective". In Karlsson, Fred; Sinnemäki, Kaius; Miestamo, Matti (Eds.), *Language complexity : typology, contact, change*. 23–43. Amsterdam/Philadelphia: John Benjamins.

Norris, John M.; Ortega, Lourdes. 2000. "Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis". *Language Learning*, *50*(3), 417–528. http://doi.org/10.1111/0023-8333.00136

Norris, John M.; Ortega, Lourdes. 2009. "Towards an organic approach to investigating CAF in instructed SLA: The case of complexity". *Applied Linguistics*, *30*(4), 555–578. http://doi.org/10.1093/applin/amp044

Ortega, Lourdes. 2003. "Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing". *Applied Linguistics*, *24*(4), 492–518. http://doi.org/10.1093/applin/24.4.492

Ortega, Lourdes. 2012. "Interlanguage complexity: A construct in search for theoretical renewal". In Kortmann, Bernd; Szmrecsanyi, Benedikt (Eds.), *Linguistic complexity: second language acquisition, indigenization, contact*. 127–155. Berlin: De Gruyter.

Ortega, Lourdes. 2015. "Syntactic complexity in L2 writing: Progress and expansion". *Journal of Second Language Writing*, *29*, 82–94. http://doi.org/10.1016/j.jslw.2015.06.008

Pallotti, Gabriele. 2009. "CAF: Defining, refining and differentiating constructs". *Applied Linguistics*, *30*(4), 590–601. http://doi.org/10.1093/applin/amp045

Pallotti, Gabriele. 2015. "A simple view of linguistic complexity". *Second Language Research*, *31*(1), 117–134.

Perneger, Thomas V. 1998. "What’s wrong with Bonferroni adjustments". *BMJ : British Medical Journal*, *316*(7139), 1236–1238.

Pienemann, Manfred. 2007. "Variation and dynamic systems in SLA". *Bilingualism: Language and Cognition*, *10*(1), 43. http://doi.org/10.1017/S1366728906002793

Pienemann, Manfred; Lenzing, Anke. 2015. "Processabiltiy Theory". In Van Patten, Bill; Williams, Jessica (Eds.), *Theories in second language acquisition: An introduction*. (2nd edition).159–179. New York: Routledge.

Polio, Charlene. 1997. "Measures of linguistic accuracy in second language writing research". *Language Learning*, *47*(1), 101–143.

Polio, Charlene; Shea, Mark C. 2014. "An investigation into current measures of linguistic accuracy in second language writing research". *Journal of Second Language Writing*, *26*, 10–27. http://doi.org/10.1016/j.jslw.2014.09.003

Ravid, Dorit; Tolchinsky, Liliana. 2002. "Developing linguistic literacy: a comprehensive model". *Journal of Child Language*, *29*, 417–447.

Robinson, Peter. 2001. "Task complexity, cognitive resources, and syllabus design: a triadic framework for examining task influences on SLA". In Robinson, Peter (Ed.), *Cognition and second language instruction*. 287–318. Cambridge: Cambridge University Press.

Robinson, Peter. 2003. "The Cognition Hypothesis, task design, and adult task-based language learning". *University of Hawai’I Second Langauge Studies Paper 21 (2)*. Retrieved from https://scholarspace.manoa.hawaii.edu/handle/10125/40656

Robinson, Peter. 2011a. "Second language task complexity, the cognition hypothesis, language learning, and performance". In *Second language task complexity : Researching the Cognition Hypothesis of language learning and performance*. 3–37. Amsterdam/Philadelphia: John Benjamins.

Robinson, Peter. 2011b. *Second language task complexity : Researching the Cognition Hypothesis of language learning and performance*. Amsterdam/Philadelphia: John Benjamins.

Robinson, Peter; Gilabert, Roger. 2007. "Task complexity, the Cognition Hypothesis and second language learning and performance". *IRAL - International Review of Applied Linguistics in Language Teaching*, *45*(3). http://doi.org/10.1515/iral.2007.007

Roca de Larios, Julio; Manchón, Rosa M.; Murphy, Liz. 2006. "Generating text in native and foreign language writing: A temporal analysis of problem solving formulation processes". *The Modern Language Journal*, *90*(1), 100–114. http://doi.org/10.1111/j.1540-4781.2006.00387.x

Schmid, Monika S.; Verspoor, Marjolijn H.; MacWhinney, Brian. 2011. "Coding and extracting data". In Verspoor, Marjolijn H.; De Bot, Kees; Lowie, Wander (Eds.), *A* *dynamic approach to second language development: methods and techniques*. 39–54. Amsterdam/Philadelphia: John Benjamins.

Schmidt, Richard. 1991. "Psychological mechanisms underlying second language fluency". *University of Hawai'i Working Papers in English as a Second Language 10 (2)*. Retrieved from https://scholarspace.manoa.hawaii.edu/handle/10125/38649

Schoonen, Rob; Snellings, Patrick; Stevenson, Marie; Van Gelderen, Amos. 2009. "Towards a blueprint of the foreign language writer: linguistic and cognitive demands of foreign language writing". In Manchón, Rosa (Ed.), *Writing in foreign language contexts: learning, teaching, and research*. 77–101. Bristol: Multilingual Matters.

Shaw, Philip; Liu, Eric Ting-Kun. 1998. "What develops in the development of second-language writing?". *Applied Linguistics*, *19*(2), 225–54. http://doi.org/10.1093/applin/19.2.225

Skehan, Peter. 1996. "A framework for the implementation of task-based instruction". *Applied Linguistics*, *17*(1), 38–62. http://doi.org/10.1093/applin/17.1.38

Skehan, Peter. 1998. *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, Peter. 2003. "Task-based instruction". *Language Teaching*, *36*(01), 1–14. http://doi.org/10.1017/S026144480200188X

Skehan, Peter. 2009. "Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis". *Applied Linguistics*, *30*(4), 510–532. http://doi.org/10.1093/applin/amp047

Skehan, Peter. 2014a. "Limited attentional capacity, second language performance, and task-based pedagogy". In Skehan, Peter (Ed.), *Processing perspectives on task performance*. Amsterdam/Philadelphia: John Benjamins.

Skehan, Peter. 2014b. "The context for researching a processing perspective on task performance". In Skehan, Peter (Ed.), *Processing perspectives on task performance*. Amsterdam/Philadelphia: John Benjamins.

Skehan, Peter; Foster, Pauline. 1997. "Task type and task processing conditions as influences on foreign language performance". *Language Teaching Research*, *1*(3), 185–211. http://doi.org/10.1177/136216889700100302

Skehan, Peter; Foster, Pauline. 2001. "Cognition and tasks". In Robinson, Peter (Ed.), *Cognition and second language instruction*. 183–205. Cambridge: Cambridge University Press.

Slobin, Dan I. 1996. "From 'thought and language' to 'thinking for speaking'". In Gumperz, John J.; Levinson, Stephen C. (Eds.), *Rethinking linguistic relativity*. 70–96. Cambridge: Cambridge University Press.

Socher, Richard; Bauer, John; Manning, Christopher D.; others. 2013. "Parsing with compositional vector grammars". In *Proceedings of the 51st Annual Meeting of the*

Association for Computational Linguistics (Volume 1: Long Papers). Vol. 1455–465. Retrieved from http://aclanthology.info/pdf/P/P13/P13-1045.pdf

Spoelman, Marianne; Verspoor, Marjolijn H. 2010. "Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of finnish". *Applied Linguistics*, *31*(4), 532–553. http://doi.org/10.1093/applin/amq001

Thelen, Esther; Smith, Linda B. 1994. *A dynamic systems approach to the development of cognition and action*. Cambridge, Massachusetts: MIT Press.

Thomas, Margaret. 1994. "Assessment of L2 proficiency in second language acquisition research". *Language Learning*, *44*(2), 307.

Towell, Richard. 2012. "Complexity, accuracy and fluency from the perspective of psycholinguisitc second language acquisition research". In Housen, Alex; Kuiken, Folkert; Vedder, Ineke (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. 47–70. Amsterdam/Philadelphia: John Benjamins.

Verspoor, Marjolijn H.; Lowie, Wander; Van Dijk, Marijn. 2008. "Variability in second language development from a dynamic systems perspective". *The Modern Language Journal*, *92*(2), 214–231.

Weideman, Albert. 2010. "Complex systems and applied linguistics". *Southern African Linguistics and Applied Language Studies*, *27*(2), 229–233. http://doi.org/10.2989/SALALS.2009.27.2.9.872

Williams, Jessica. 2012. "The potential role(s) of writing in second language development". *Journal of Second Language Writing*, *21*(4), 321–331. http://doi.org/10.1016/j.jslw.2012.09.007

Wolfe-Quintero, Kate; Inagaki, Shunji; Kim, Hae-Young. 1998. *Second language development in writing: measures of fluency, accuracy, & complexity*. Honolulu, Hawaii: University of Hawai’i Press.

Wouter, Kusters. 2008. "Complexity in linguistic theory, language learning and language change". In Karlsson, Fred; Sinnemaki, Kaius; Miestamo, Matti (Eds.), *Language complexity : typology, contact, change*. 3–22. Amsterdam/Philadelphia: John Benjamins.

Yang, Weiwei; Lu, Xiaofei; Weigle Cushing, Sara. 2015. "Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality". *Journal of Second Language Writing*, *28*, 53–67. http://doi.org/10.1016/j.jslw.2015.02.002

Appendix

1 Info-sheet and questionnaire

The info sheet and the following questionnaire was created by following Dörneyi’s guidelines for research design in applied linguistics (2007).

Info sheet for study participation

Dear colleagues, my name is Thomas and in my Diploma Thesis I would like to analyze Proseminar & Seminar papers of students from the English department for syntactic complexity. To be able do this I rely on your help. It would be great if you could...

...fill in the attached leaner profile

...send me a copy of your linguistics Proseminar paper 1 (preferably as .doc, .docx or .odt)

AND

...send me a copy of your linguistics Seminar paper at the end of the semester

All your data will be anonymized & treated confidentially! This means there will be no way to trace back any of the data to you, your name or your matriculation number. If you have any questions about my diploma thesis or your provided data, do not hesitate to contact me per mail.
It would be great if you could send a copy of your PS linguistics paper to my mail address within the following days: a1026361@unet.univie.ac.at (Thomas WALTER)

You can keep this info sheet and I´ll send you an e-mail reminder on your university mail address in February to send me a copy of your SE linguistics paper.

Thank you very much for your help!
Thomas WALTER

(Page 1 of 2)

Learner profile questionnaire

1) General &Information:

Name:……………………………………… E-mail: ……………………………………….
Age:……… Gender:……………………...Matrikelnr.:……………………………..

2) Mother tongue(s): o German o other(s): ………………………………………………

3) Do you use English regularly in your current living situation? o yes o no

4) Years of formal English education in school:…………years

5) Upper secondary school type: o AHS o BHS o other:………………….

6) Which study program are you in?
o Bachelor o Master o Teaching (2nd subject:………………………………)

7) Semesters studied English at university level:…………semesters

8) Time spent in English speaking countries (if longer than 1 month):…………..months

9) In which semester did you write your Proseminar linguistics 1 paper?………………

10) Which writing courses did you take between PS linguistics 1 and now?

o LIU 1 o LIU 2 o EAP o EPCO o other:………………………………

11) Do you think the complexity of your writing increased between writing your PS linguistics 1 paper and now? o yes o no

12) How would you rate your writing proficiency in English?

o B2 o B2/C1 o C1 o C1/C2 o C2 o I don´t know

Please turn over ;-)

(Page 2 of 2)

13) When I'm writing, I like to try out new ways of expressing my thoughts:

Disagree o o o o o Agree

14) What other languages *except* your mother tongue(s) & English do you speak?

o none estimated proficiency level

language: ………………………………. o A1 o A2 o B2 o B1 o C1 o C2

language: ………………………………. o A1 o A2 o B2 o B1 o C1 o C2

language: ………………………………. o A1 o A2 o B2 o B1 o C1 o C2

15) What makes up complex writing for you (key words are fine)?

……………………………………………………………………………………………

……………………………………………………………………………………………

16) In which context did you use your English writing skills the most over the last year?

o at university o at work o with friends or family o other:………………….

Comments, thoughts, feedback (optional):

……………………………………………………………………………………………

…………………………………………………………………………………...……

I have read the information sheet and I agree that the above data and my two papers may be used for scientific purposes.

Date:……………………………. Signature:………………………………

Thank you very much for your help!

2 Learner background information

Table 9: Detailed sample description

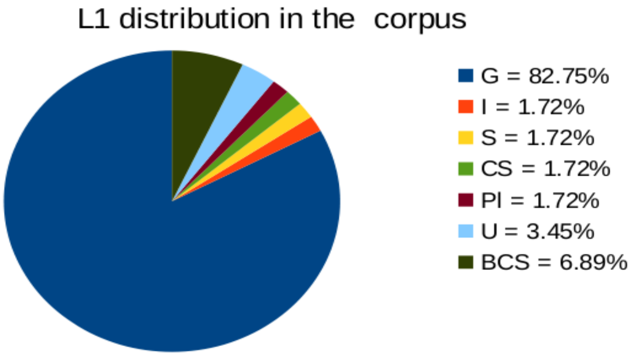
Background data for each sample (part 1/2)

Sample no.	Gender	Secondary School	L1	2nd L1	Formal English education [y]	Time at Uni [semesters]	Age [y]	Time PS to SE [semesters]	No. writing courses btw. PS and SE	PS wordcount [words]	SE wordcount [words]	PS paper extract [words]	SE paper extract [words]
1	f	AHS	G	no	7	6	27	2	3	2415	6067	1989	1992
2	f	BHS	BCS	no	9	9	23	5	4	2292	6856	1933	1986
3	f	AHS	BCS	no	8	9	22	5	2	2732	6507	2005	2006
4	f	AHS	G	no	8	9	23	4	4	2379	7823	1992	1997
5	f	AHS	G	no	8	10	23	3	1	2771	8314	1978	1977
6	m	BHS	G	no	9	7	23	2	1	2314	5061	1942	1979
7	f	BHS	G	no	10	7	22	3	2	2320	8593	1980	1983
8	f	AHS	I	no	13	7	22	3	3	2579	7396	1987	1991
9	f	AHS	G	no	12	9	24	3	2	2314	6390	1979	2002
10	f	BHS	G	no	9	11	25	6	3	2809	8300	1976	1998
11	f	BHS	G	no	9	10	25	6	4	2747	4658	1984	1984
12	f	AHS	G	no	12	9	25	7	2	2423	6521	1964	1997
13	f	AHS	G	no	12	5	21	2	2	2195	8733	1961	1997
14	f	BHS	G	no	9	11	26	8	4	2447	7571	1992	1982
15	f	AHS	G	no	8	8	22	3	3	2731	6994	1964	1987
16	f	Spain	S	no	12	6	27	2	1	2793	6096	1988	1989
17	f	Ukraine	U	no	8	5	24	3	2	1850	5939	1770	2004
18	f	BHS	G	no	9	12	27	5	4	2725	7055	1998	1986
19	f	AHS	G	no	10	9	24	2	2	2162	8847	2002	2006
20	f	BHS	G	no	10	9	24	5	4	2695	9068	1997	1994
21	f	AHS	G	no	12	6	23	3	3	2583	8976	1959	1974
22	f	AHS	G	no	8	9	22	3	4	2309	6986	1988	1995
23	f	Ukraine	U	no	8	4	28	3	2	2290	6999	1977	1986
24	f	BHS	G	no	9	5	24	2	1	2594	8065	1871	1972
25	f	AHS	Pl	G	13	7	22	3	3	1959	11171	1723	2000
26	m	BHS	G	no	13	8	26	3	2	2695	5987	1972	2004
27	f	Slovakia	CS	no	13	8	22	5	2	2606	10040	1986	2011
28	f	AHS	G	no	8	9	22	4	2	2270	6100	1879	1991
29	f	AHS	G	no	8	7	22	5	4	2222	8377	1982	1975
30	f	BHS	G	no	9	9	23	6	4	2364	5863	1985	1993
31	f	BHS	G	no	9	11	25	6	4	2549	6413	1990	2009
32	f	AHS	BCS	G	8	11	23	5	4	2997	6749	1971	1972
33	f	AHS	G	no	8	10	25	3	3	2987	7361	2004	2025
34	f	AHS	G	no	8	7	21	4	4	2665	6811	1966	1981
35	f	AHS	G	no	8	8	25	9	3	2317	8330	1984	1991
36	m	AHS	G	no	8	9	23	4	3	2437	6258	1982	2000
37	f	AHS	G	no	12	9	23	3	2	2727	6242	1991	1990
38	f	AHS	G	no	8	13	25	4	2	2369	6521	1780	2006
39	f	BHS	G	no	9	11	24	5	1	2873	6940	1993	1975
40	f	AHS	G	no	8	9	23	4	3	2743	7222	1976	2001
41	m	AHS	BCS	no	5	11	29	14	1	9103	6557	1994	2022
42	f	AHS	G	no	9	7	21	4	4	2425	6984	1973	1934
43	f	AHS	G	no	8	9	23	2	3	2099	5870	1959	1979
44	m	AHS	G	no	8	12	26	7	2	3159	6240	1991	2001

Background data for each sample (part 2/2)

Sample	Gender	Secondary School	L1	2^{nd} L1	Formal English education [y]	Time at Uni [semesters]	Age [y]	Time PS to SE [semesters]	No. writing courses btw. PS and SE	PS wordcount [words]	SE wordcount [words]	PS paper extract [words]	SE paper extract [words]
45	f	AHS	G	no	10	5	20	2	2	2435	10666	1997	1992
46	f	BHS	G	no	9	8	22	4	3	2356	9163	1988	1992
47	f	BHS	G	no	13	7	22	3	3	2702	6961	1950	1979
48	m	AHS	G	no	8	10	32	4	3	2402	6959	1994	2002
49	m	BHS	G	no	12	14	30	8	2	2998	6826	1984	1982
50	f	AHS	G	no	8	9	24	4	3	2749	7366	1996	1998
51	f	AHS	G	no	8	9	47	2	1	2744	8893	2009	1989
52	f	AHS	G	no	8	9	23	6	3	2593	8965	1968	1981
53	m	AHS	G	no	8	7	23	2	1	2236	6401	2011	1980
54	f	BHS	G	no	10	9	25	5	2	2497	6879	1964	1997
55	m	AHS	G	no	8	10	24	5	3	2739	6024	1964	1996
56	f	AHS	G	no	8	7	22	3	1	2444	6980	1789	1975
57	f	BHS	G	no	13	7	22	2	2	2320	5795	2003	1995
58	m	BHS	G	no	11	13	31	7	4	2425	6238	1953	1991
				Sum	-	-	-	-	-	152675	419967	113827	115473
				Mean	9.4	8.6	24.4	4.3	2.7	2632.3	7240.8	1962.5	1990.9
				STDev	1.9	2.1	3.9	2.2	1.1	903.5	1317.6	60.1	14.9
				STDerr.	0.25	0.28	0.51	0.29	0.14	118.64	173.01	7.9	2.0
				Median	9.0	9.0	23.0	4.0	3.0	2472.0	6960.0	1982.0	1992.0
				Min	5	4	20	2	1	1850	4658	1723.0	1934.0
				Max	13	14	47	14	4	9103	11171	2011.0	2025.0

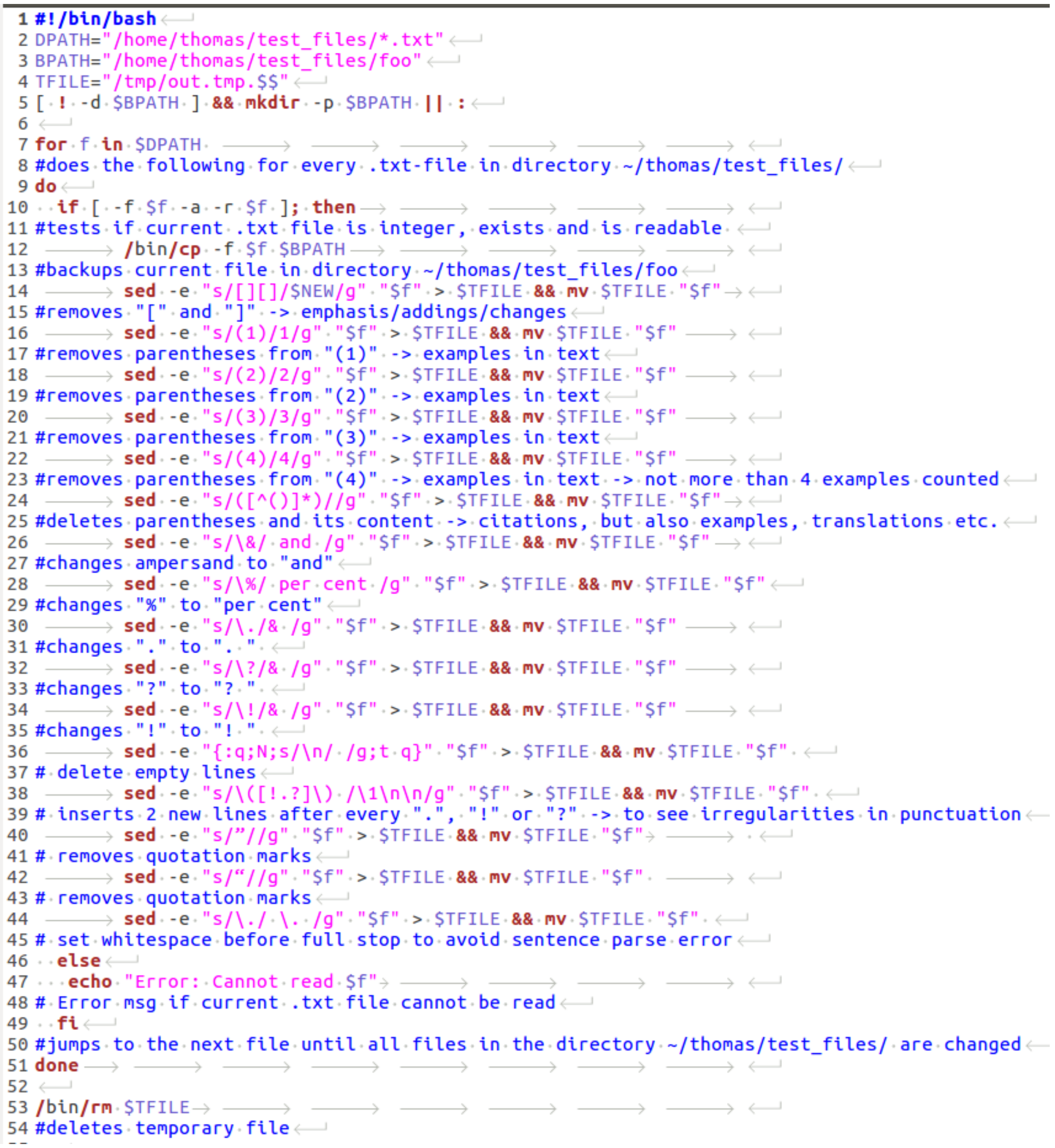
	Samples	**Bilingual (all G)**
G=German	48	no
I=Italian	1	no
S=Spanish	1	no
CS=Slovakian	1	no
Pl =Polish	1	yes
U=Ukrainian	2	no
BCS=Bosnian, Croatian or Serbian	4	1 of 4



Gender	**Samples**	**[%]**
female	48	82.76
male	10	17.24

School type	**Samples**	**[%]**
AHS	35	60.34
BHS	19	32.76
other	4	6.90

3 Script for editing the input files



```bash
#!/bin/bash
DPATH="/home/thomas/test_files/*.txt"
BPATH="/home/thomas/test_files/foo"
TFILE="/tmp/out.tmp.$$"
[ !  -d $BPATH ] && mkdir -p $BPATH || :

for f in $DPATH
#does the following for every .txt-file in directory ~/thomas/test_files/
do
  if [ -f $f -a -r $f ]; then
#tests if current .txt file is integer, exists and is readable
	/bin/cp -f $f $BPATH
#backups current file in directory ~/thomas/test_files/foo
	sed -e "s/[][]/$NEW/g" "$f" > $TFILE && mv $TFILE "$f"
#removes "[" and "]" -> emphasis/addings/changes
	sed -e "s/(1)/1/g" "$f" > $TFILE && mv $TFILE "$f"
#removes parentheses from "(1)" -> examples in text
	sed -e "s/(2)/2/g" "$f" > $TFILE && mv $TFILE "$f"
#removes parentheses from "(2)" -> examples in text
	sed -e "s/(3)/3/g" "$f" > $TFILE && mv $TFILE "$f"
#removes parentheses from "(3)" -> examples in text
	sed -e "s/(4)/4/g" "$f" > $TFILE && mv $TFILE "$f"
#removes parentheses from "(4)" -> examples in text -> not more than 4 examples counted
	sed -e "s/([^()]*)//g" "$f" > $TFILE && mv $TFILE "$f"
#deletes parentheses and its content -> citations, but also examples, translations etc.
	sed -e "s/\&/ and /g" "$f" > $TFILE && mv $TFILE "$f"
#changes ampersand to "and"
	sed -e "s/\%/ per cent /g" "$f" > $TFILE && mv $TFILE "$f"
#changes "%" to "per cent"
	sed -e "s/\./& /g" "$f" > $TFILE && mv $TFILE "$f"
#changes "." to ". "
	sed -e "s/\?/& /g" "$f" > $TFILE && mv $TFILE "$f"
#changes "?" to "? "
	sed -e "s/\!/& /g" "$f" > $TFILE && mv $TFILE "$f"
#changes "!" to "! "
	sed -e "{:q;N;s/\n/ /g;t q}" "$f" > $TFILE && mv $TFILE "$f"
# delete empty lines
	sed -e "s/\([!.?]\) /\1\n\n/g" "$f" > $TFILE && mv $TFILE "$f"
# inserts 2 new lines after every ".", "!" or "?" -> to see irregularities in punctuation
	sed -e "s/”//g" "$f" > $TFILE && mv $TFILE "$f"
# removes quotation marks
	sed -e "s/“//g" "$f" > $TFILE && mv $TFILE "$f"
# removes quotation marks
	sed -e "s/\./ \. /g" "$f" > $TFILE && mv $TFILE "$f"
# set whitespace before full stop to avoid sentence parse error
  else
   echo "Error: Cannot read $f"
# Error msg if current .txt file cannot be read
  fi
#jumps to the next file until all files in the directory ~/thomas/test_files/ are changed
done

/bin/rm $TFILE
#deletes temporary file
```

All papers were converted to .txt-files. Headings and quotes longer than two lines were manually removed. Then the above script was executed in the terminal of Ubuntu 16.04. Afterwards all files have been manually cut to approximately 2000 words and special characters that were not removed by the script (e.g. “...”, “ð”, “æ”, etc.) were adjusted or deleted (for data manipulations see the following examples in the appendix). Furthermore, numbers with commas (e.g. “...in chapter 3.1 the results...” or “…the first groups shows a mean of 34.5 per cent…”) were adjusted to the nearest integer value, because commas are detected as the end of a sentence by the L2SCA. The blue lines beginning with “#” mark comments in the script and explain what the previous command modified in the text file. The file was written with the help and the partial use of code from the following sources: Barnett 2015; Goyvaerts 2016; Mesibov 2014.

4 Example paragraphs before and after modifications

The following pages show example paragraphs of texts before and after the script has been run and the manual modifications have been applied. The formatting (except the font style) has been taken one to one from the text files, including spacing.

Example from PS paper, sample 17 before modifications:

The point of view on errors has frequently changed in recent years and in CLT “making mistakes is […] considered to be a necessary part of a […] learner’s progress towards mastery of the language” (Revell 1979: 8). A student can only fully understand a newly learned word, phrase, grammatical item, etc. by using it in different contexts. That way, the language learner experiences the possibilities of its use and by making mistakes he/she gets to know the limits of application of structures (E.g.: of vocabulary) (Ibid.).

Example from PS paper, sample 17 after modifications:

The point of view on errors has frequently changed in recent years and in CLT making mistakes is considered to be a necessary part of a learner’s progress towards mastery of the language .

A student can only fully understand a newly learned word, phrase, grammatical item by using it in different contexts .

That way, the language learner experiences the possibilities of its use and by making mistakes he/she gets to know the limits of application of structures .

Example from SE paper, sample 13, before modifications:

2. Power, ideology and textbook discourse
Since the origin of critical linguistics in the 1970s and the subsequent development of critical discourse analysis (henceforward CDA), experts in the field have been discussing the close conjunction between language, ideology and “complex differential power relationships” omnipresent in daily social and political interactions (Mesthrie 2000: 316) and have provided frameworks for the successful application of this knowledge to various contexts (Zotzmann & O’Regan 2016: 115).

Example from SE paper, sample 13, after modifications:

Since the origin of critical linguistics in the 1970s and the subsequent development of critical discourse analysis , experts in the field have been discussing the close conjunction between language, ideology and complex differential power relationships omnipresent in daily social and political interactions and have provided frameworks for the successful application of this knowledge to various contexts .

Example from SE paper, sample 26, before modifications:

All of the three sentences presenting friendliness as a typical character trait of the members of the class ‘tiger’. The only distinction that could be made is by saying that (1) and (3) refer to the whole class whereas (2) refers to each member of the respective class.

Example from SE paper, sample 26, after modifications:

All of the three sentences presenting friendliness as a typical character trait of the members of the class ‘tiger’ .

The only distinction that could be made is by saying that 1 and 3 refer to the whole class whereas 2 refers to each member of the respective class .

Example from PS paper, sample 9, before modifications:

Further, they state that this region, like the Upper North, retained the r and the language in this area contains the [æ] in fast, ask, grass, etc. According to Baugh & Cable (2004:380), the Middle Atlantic region is a sub-region of the Lower North and its distinction in speech is the “unrounded vowel in forest and in hot, the [e] of egg in care, Mary and merry” and it also merges the sounds “[o] and [ɔ] before [r] in four and forty”.

Example from PS paper, sample 9, after modifications:

Further, they state that this region, like the Upper North, retained the r and the language in this area contains the ae in fast, ask, grass, etc .

According to Baugh and Cable , the Middle Atlantic region is a sub-region of the Lower North and its distinction in speech is the unrounded vowel in forest and in hot, the e of egg in care, Mary and merry and it also merges the sounds o and o before r in four and forty .

5 Texts used for L2SCA comparison

Sample 1, PS paper text

1. Anyone who has learned a foreign or second language at some point in life has probably experienced certain changes after some time without practicing the language.
2. Some words might be forgotten and speech production will become more hesitant; in other words, certain language skills deteriorate.
3. Maintaining skills in a foreign or second language requires a lot of effort when the everyday language environment does not provide reinforcement.
4. Under certain circumstances, not only second language knowledge can decrease, but first language competence can also undergo changes.
5. Variables influencing this type of deterioration are often connected to a relocation to a different country.
6. This paper examines the situation of native speakers of a language who migrate to a country with a different linguistic environment and experience a process resulting in the decrease of the ability to produce and receive the language of origin.
7. This phenomenon is referred to as first language attrition .
8. The aim of this paper is to examine the extralinguistic variables that play an important role in the occurrence of first language attrition.
9. The focus of the paper lies on first language attrition in sequential bilinguals in a L2 environment.
10. This paper consists of two main parts: the first section attempts to define first language attrition, and the second section introduces the extralinguistic factors which have been shown to play an important rolein first language attrition.
11. Even though research on language attrition has substantially grown over the last three decades, many questions have still not been answered to a sufficient extent.
12. While some studies suggest that the occurrence of language attrition in an L1 might not even be possible, other researchers focus on the reasons for its occurrence, and are, therefore, convinced that the phenomenon does exist .
13. It is, hence, not surprising that agreement on a generally applicable definition of the term could also not be reached.
14. This is,moreover, connected to the fact that the research field of language attrition is very diverse and includes, inter alia, psycholonguistic, neurolinguistic and sociolinguistic approaches whose researchers define first language attrition according to the adopted approach.
15. According to a certain researchers , first language attrition can occur without the influence of another language as demonstrated in the desert island situation, in which a person’s exposure to an L1 is effectively ceased .
16. This is often the case when native speakers migrate to a country with a different dominant language than their L1.
17. These sequential bilinguals acquire and use an L2; as a result, changes in the first language system can occur.
18. In this context, Schmid uses language attrition to refer to the phenomena of L1 change and L2 interferences .
19. Changes of this type can occur on different levels, for instance, the lexical,morphological or pragmatic level.
20. Furthermore, the reception as well as the production of the L1 can undergo changes resulting in hesitant speech when searching for words, or switching between the two languages .
21. For the purpose of the present discussion the case of drastic changes in the linguistic environment of speakers acquiring a L2 as a result of emigration and, consequently, becoming sequential bilinguals is assumed.
22. The most important factors influencing the emergence of first language attrition in the described situation will be examined in the next section.
23. Studies show that many different variables influence first language attrition.
24. The paper will focus on four factors which are considered to be highly relevant.

25. Even though the factors will be examined individually, many of them correlate with each other.
26. This connection between the single factors must be kept in mind when doing research in the area.
27. The extralinguistic variables analyzed in this section depend on a speaker’s biography as well as personal factors.
28. More precisely, this section focuses on age, L1 input and output, level of education and attitude.
29. Among the variables influencing first language attrition, age appears to be one of the most important factors.
30. Research has shown that first language attrition occurs in children as well as in adult speakers of a language.
31. The development of the phenomenon, however, does diverge significantly between children and adults .
32. Studies have shown that the first language attrition process is, compared to adults, very severe in children who find themselves in a new linguistic environment .
33. This might be connected to the fact that linguistic and cognitive competences in children are less developed and, therefore, they are more vulnerable to first language attrition .
34. Studies investigating the age at onset of attrition identified an age effect comparing participants exposed to a new linguistic environment pre-puberty and another group of participants experiencing a change in language at a post-puberty stage .
35. The study showed that children at a pre-puberty stage were more affected by first language attrition.
36. Language attrition can even result in a deep decay of the language system when the child leaves the country of origin at a very early stage and the contact with the L1 is completely interrupted, for instance, in the case of adopted children .
37. Furthermore, Pallier studied the case of adopted Korean children whose first language attrition resulted in the incapability of even recognizing their L1.
38. Pallier claim that the L1 is completely substituted by the language of the new environment because there is no need to maintain the language of origin .
39. The results of studies undertaken with children as subjects strongly contrast with the research on first language attrition in adults.
40. According to those studies, adults usually have developed a very stable linguistic proficiency in their L1, which makes attrition emerge only in a few sub-systems of the L1 the lexicon or phonological ability .
41. Other important areas of the L1 are hardly influenced by attrition even if the speaker has not been exposed to or used it in a very long time ; therefore, the proficiency level remains astonishingly high .
42. These results suggest that there must be some kind of age limit conditioning the occurrence of first language attrition.
43. Even though researchers do not agree at what age language attrition is likely to emerge, the phase between childhood and adolescence seems to represent an important threshold .
44. Knowledge in the human brain has to be activated regularly in order to prevent decay because the more certain information is triggered the easier it gets to remember and reuse it.
45. On the other hand, knowledge that has not been reactivated over a very long period is much more difficult to recall .
46. Therefore, it may be reasonable to expect that, besides the age factor, the lack of input and output of an L1 plays an important role in first language attrition.
47. Seliger and Vago acknowledge the reduced usage of an L1 as one of the most important factors in the occurrence of first language attrition .
48. The role of lack of contact for first language attrition is supported by direct evidence as well as indirect evidence from studies that expected more attrition, but resulted in little evidence for it and attribute the findings to the fact that the participants were still substantially exposed to the L1 .
49. However, the number of studies supporting this assumption is quite small.
50. According to Köpke and Schmid this might be related to the fact that measuring the frequency of use of a language depends most of the time on self-reports on questionnaires .
51. Moreover, research participants’ definitions of frequency of use might differ greatly from one another and instances, like writing into a diary, swearing, counting, might be excluded entirely even if they occur on a daily basis .
52. In order to be able to measure frequency of use more sufficiently, different approaches should be taken into account.
53. Schmid proposes the differentiation between interactive L1 use , non-interactive exposure: books and the media and the language of thought and emotion .
54. Furthermore, she notes that passive exposure and personal usage of an L1 might outweigh active production including conversation, quantitatively.
55. The latter might however play a more important role in the preservation of the L1 and, consequently, the prevention of first language attrition.
56. Therefore, these approaches need to be considered in future research .
57. Even though the level of education might be connected to other factors such as age and is difficult to define and operationalize, it appears to be another important variable for first language attrition.
58. One of the first studies to find a correlation between the level of education and attrition was a pilot study by Jaspaert and Kroon .
59. They conducted a variety of tests on Italian immigrants in the Netherlands and the educational level appeared to be the most important factor for, particularly in two types of test, the text editing and the vocabulary test.
60. As a result, Jaspaert and Kroon concluded that higher educated participants performed better in certain tests, and are, therefore, more capable of preserving their L1 in a L2 environment than participants with lower education .
61. Furthermore, Köpke and Schmid confirm that more higly educated speakers dispose of more profound knowledge about language; as a result, the probability of interference decreases .
62. The results might also be connected to the social status of the participants and possibly correlate with a higher salary.
63. This, again, could result in better tools for language maintenance, for instance, more frequent travel to the country of origin and hence contact with their L1 , which was earlier acknowledged as an important factor for the prevention of first language attrition.
64. Another relevant aspect depending on a person’s educational background is literacy in the L1.
65. Knowing how to write and/or read has been established as another important facet in the prevention of first language attrition because it helps to stabilize the language in the brain by adding a new network .
66. This provides further evidence for the argument presented in the subsection concerning the age factor why first language attrition occurs frequently and with more serious effects in children than in adults.
67. It is probable that older children and adults are literate in their first language; therefore, first language attrition affects children who leave their country of origin before concluding the process of becoming literate more strongly.
68. The importance of a bilingual’s attitude towards his/her L1, especially in a migrant situation, is considered to be one of the most crucial extralinguistic factors for the development of a language.
69. Accordingly, speakers with a negative attitude towards their L1 are more likely to be affected by first language attrition.

70. In contrast, first language attrition is expected to occur less frequently in speakers who show a positive attitude towards their L1 .
71. Like many of the other affective variables, language attitude is equally difficult to measure.
72. On the one hand, the attitude adopted towards a language is dynamic, which means that the attitude of a speaker towards a L1 can change over time.
73. Nostalgia or an emotional distance from the language of origin might, for instance, influence the motivation for using the L1 .
74. On the other hand, attitude is a unit that cannot simply be observed and recorded, but must be elicited from spoken data.
75. This leads to another difficulty in the investigation because participants who self-report tend to answer in a way they think is appropriate or expected; therefore, the information might not be reliable and a different method needs to be developed .
76. Nevertheless, some recent studies have successfully determined a connection between attitude and first language attrition using a new methodology.
77. Prescher collected data by means of thorough interviews with German migrants in the Netherlands.
78. The results confirmed a connection between language development, attitude and identity.
79. The positive attitude of participants in Prescher's study towards their L1 and culture increased in accordance with the duration of residency in the L2 environment.
80. Consequently, the participants' desire to maintain their L1 increased and the contact with the L1 was reestablished .
81. Another recent study acknowledges the importance of attitude in the context of first language attrition; however, the research also demonstrated that a positive attitude does not guarantee the prevention of first language attrition.
82. Nevertheless, a positive attitude towards the L1 increases the motivation for the maintenance of the language.

Sample 1, SE paper text

1. The phenomenon of multilingualism is not a recent occurrence.
2. It has already been part of the linguistic reality in the past because of certain societal, economic and political developments which caused migration and made communication betweendifferent cultural and linguistic groups necessary .
3. Current multilingualism has different caracteristics and is connected to concepts such as globalization and mobility .
4. The reasons for multilingualism to emerge, however, do not have to be environmental but can also be personal .
5. In linguistics, monolinguals have been and are to this day often taken as the norm for reference when doing research, even though the majority of people all over the world speaks more than one language, and could be defined as being multilingual in the widest sense.
6. Hence, multilingualism should be taken as the conception that serves as the norm for linguistic research .
7. This is the reason for which the focus of this papers lies on multilingualism and tries to investigate its connection to a person's ability to learn a foreign language, a concept also defined as language aptitude.
8. Language aptitude or, to be more specific, foreign language aptitude generally refers to the talent of a person regarding foreign language learning; it is, however, also used in the context of second language learning .
9. Many disciplines have taken an interest in language aptitude in the course of the last years, for instance applied linguistics; and the different fields influence each other.
10. More recent research in the field of cognitive sciences, more specifically cognitive psychology and neuroscience which apply brain imaging, has also significantly contributed to an increase of interest in the topic of language aptitude.
11. The multitude of disciplines doing research on language aptitude has helped the development of the topic and even changed the concept of language aptitude .
12. In this regard, it might be useful to provide an overview of possible definitions of the concept.
13. One of the first researchers to work in the field of language aptitude was Carroll who defined the term as follows: an individual's initial state of readiness and capacity or learning a foreign language, and probable facility in doing so given the presence of motivation and opportunity .
14. Carroll also insisted on differentiating between the terms aptitude and achievement.
15. Moreover, he drew a distinction between language aptitude and general intelligence, because although the two are connected, one is not a prerequisite for the other .
16. Language aptitude was originally seen as a rather stable, if not innate ability of an individual to learn a language.
17. As has already been mentioned, the conceptualization of language aptitude has been adapted because of more recent developments and discoveries in research on the concept.
18. While scientists in the early stages of language aptitude research define the concept as rather stable and innate, newer research resulted in a more dynamic definition of the term language aptitude.
19. One of the scientists challenging the original conceptualization was Sternberg who sees language aptitude as something strongly influenced by external factors and depending on the language learning experiences ofan individual , hence, language aptitude is seen as a more dynamic concept.
20. Skehan shares Sternberg's previously mentioned point of view and defines aptitude as being characterized by two criteria.
21. On the one hand, there is a person's natural ability to process language, and on the other hand this innate ability is complemented by certain external factors.
22. Other scientists such as Dörnyei take it even a step further and argue that the concept of language aptitude in itself does not exist.
23. According to Dörnyei, a variety of different cognitive factors play a role in a person's ability for learning a foreign language and not only a talent for language learning, a way language aptitude is also referred to mainly by laypeople.
24. The purpose of the present paper is to investigate the influence of different factors, but particularly a person's previous experience in language learning on language aptitude because it is assumed that language aptitude is in fact a dynamic concept that fluctuates.
25. More specifically, the potential interplay between multilingualism and language aptitude is examined.
26. Before starting the discussion on said connection, it might be useful to provide a definition of multilingualism relevant to this paper.
27. Even though most people have a concept in mind when hearing multilingualism, its definition proves to be a more difficult endeavor.
28. The term multilingualism includes a wide range of meanings.
29. According to Stavans and Hoffmann bilingualism originally included contexts in which more than one language was involved, but mainly the contact between two languages.
30. It also comprised, however, situations of trilingualism or multilingualism.

31. In more recent research adopting a perspective that goes beyond two languages bilingualism is considered to be a subcategory of multilingualism .
32. Stavans and Hoffmann suggest to approach trilingualism and multilingualism separately from bilingualism.
33. At the basis of this paper, lies a more liberal definition of multilingualism only referring to the number of languages involved, and not to the manner in which the languages have been acquired.
34. This paper is based on a definition of multilingualism which defines being multilingual as knowing more than two languages, hence, a minimum of three or more languages.
35. An important aspect to consider when investigating the connection between language aptitude and multilingualism is the linguistic background of multilinguals.
36. In this regard it is interesting to investigate the way language is processed by multilingual speakers.
37. In the past two decades, researchers have focused on developing language processing models suitable for multilinguals.
38. At the basis of this more recent interest is the assumption, that the different thought-processes involved in a multilingual's brain vary a lot from the processes involved in bilingual speakers.
39. Herdina and Jessner developed a Dynamic Model of Multilingualism which presents the language systems in existence in a multilingual individual as being influenced on a social, psycholinguistic and linguistic level .
40. Multilingual proficiency is regarded as the result of an interplay between psycholinguistic systems, cross-linguistic elements and the multilingualism factor .
41. Another element that distinguishes multilinguals from monolinguals, but also from bilinguals, is the fact that because of the different language systems involved in a multilingual's mind, the processes become more complex with the number of languages, their structure, typology, usage and availability .
42. Hence, a high degree of control of the language systems is necessary.
43. Results in brain imaging in children has shown that adaptations in a multilingual's brain can be observed.
44. The region in the brain concerned by structural changes is responsible for enhancing attention functions deployed in order to monitor and control the different languages being spoken .
45. Della Rosa confirms that it is possible for multilinguals to have cognitive advantages, such as a good ability to focus attention .
46. Another element that is investigated in research about language aptitude is the memory capacity of language learners.
47. Regarding multilinguals and their working and short-term memory, a study by Biedroń and Szczepaniak has shown that multilinguals have better overall memory capacities when tested on linguistic material than other, non-mulitlingual participants.
48. The scholars, however, also conclude that it cannot be determined whether the high memory abilities developed because of their achievements or if their language experience has caused them to have an advantage in certain memory tasks.
49. They consider it probable that both stand in relation to each other .
50. Connected to the linguistic background of a person is the language learning experience a multilingual has gained in the course of his or her life.
51. Herdina and Jessner argue that monolinguals, bilinguals and multilinguals employ different strategies when learning a foreign language.
52. Therefore, mutilinguals seem to be more effective learners of languages, hence, they do it with more ease than for instance monolinguals .
53. Some research on the topic claims that the reason for the higher facility of learning new languages can be traced to higher metalinguistic awareness that is displayed by multilinguals.
54. According to Thompson, the multitude of languages of a multilingual learner influences a person's construction of the internal linguistic system and, therefore, it also affects a learner's language aptitude .
55. According to Nayak, the more experienced a language learner is, and multilinguals could be categorized as having said experience, the easier it is for them to adapt the learning strategies to an unknown language they try to learn.
56. This is linked to the fact that an increased awareness about the existence of different language systems and their characteristics increases the ability of adaptation .
57. However, most scholars would agree that conscious training is a prerequisite in order for metalinguistic awareness to develop and, hence, to have an advantage being multilingual .
58. Since training is regarded as a precondition for the development of metalinguistic awareness, a study on behavioral discrimination abilities has been conducted regarding potential benefits or disadvantages of multilinguals in the area of sound perception by Tremblay and Sabourin .
59. The study tested the ability of mono-, bi- and multilinguals to distinguish a nonnative contrast behaviorally in the pre-training and post-training phase.
60. The study has shown that the more experienced language learners are, the better they also are at learning languages .
61. Not only linguistic factors play a role in language learning.
62. Personal factors can also be significant in this regard.
63. Since the paper has a restricted research focus, the major individual factors considered in this paper are motivation, self-assessed proficiency scores and an element called multilinguality factor which varies considerably between one individual and the next.
64. However, there is a vast number of personal factors not taken into consideration in this paper, for instance learning strategies or learning styles that might also play a role as has been discussed in the course of the seminar by other colleagues.
65. In individual differences research, motivation is seen as one of the most important factors for successful language learning .
66. Carroll became aware of the importance of certain personal factors that influence language aptitude early on.
67. He assumed that certain factors that are not directly associated with languages, but rather with an individual's attitude towards learning languages, like motivation can be crucial for language learners.
68. Therefore, he claims that a person with low language aptitude can still have high language achievement results if the motivation of said learner is high.
69. Other scientists such as Dörnyei and Skehan even assume that motivation influences the learning behavior of a learner and, therefore, also the language aptitude.
70. They suggest a dynamic model for motivation.
71. Other scientists take an individual's personal attitude towards the foreign language and the culture connected to it into consideration, and regard the personal motivation of a language learner as crucial for successful foreign language learning .
72. Connected to a person's personal motivation for learning a language is the selfassessed proficiency of languages already existing in a person's linguistic repertoire, since previous success in language learning might influence a person's motivation tremendously.
73. Research on the influence of a person's proficiency on future language learning is usually only investigated if learners are very proficient in the languages .

74. Other scholars have, however, determined that already minor knowledge in a language and low proficiency in a nonnative language can have a positive effect on future language learning .
75. The proficiency of a person can either be assessed by means of a proficiency test for the language underinvestigation.
76. Another means of assessing the proficiency of a person in a language is often applied through questionnaires in which participants self-assess their proficiency in a language or their languages.
77. The questionnaire about the linguistic background knowledge given to participants for this study also included a section regarding the proficiency of the languages since having a certain proficiency level was no prerequisite for the study at hand.
78. A high self-assessed proficiency could also be an indicator for the positive learning experiences of a multilingual which might consequently have an effect on language aptitude .
79. The multilinguality score calculated for participants in this study is a factor included in this paper that considers linguistic factors as well as non-linguistic, or rather personal elements that play a potential role for language aptitude and it is a strongly individual factor.

Sample 28, PS paper text

1. Through the course of history, the area of New York City has developed a characteristic pronunciation and dialect of North American English, certainly as recognizable as the city itself.
2. Especially within the United States, citizens can be easily identified as New Yorkers, as the New York City dialect has evolved to a trademark often presented comically on television or radio .
3. However, New York City English cannot be perceived as one single dialect, as the city consists of an enormous amount of different social, national and ethnic groups.
4. Ethnicity controls the use of language in various ways, as different ethnicities or nations often speak a certain language.
5. However, in prevailing languages such as English, ethnic identities also affect variation within a dialect , even though the accurate amount on how much certain ethnic groups have contributed to a certain dialect is not easily identified , as dialects as such are influenced by a vast number of social, economic and linguistic factors.
6. Nevertheless, a number of recognizable ethnic dialects have developed through the course of history in North American English, such as Italian English, Jewish English, Irish English, German English, Puerto Rican English, Chicano English, American Indian English, Vietnamese English, and, of course, AAVE .
7. The purpose of this paper is, consequently, to give not only an overview of the peculiarities of the pronunciation of English in New York City, but also to analyze ethnic influences on phonetic features, such as rhoticity and vowel phonologies.
8. New York City English is traditionally the most famous nonrhotic dialect of North American English .
9. As the city is situated on the East Coast, its nonrhoticity can also be attributed to British settlement, maintaining linguistic parallels with their mother city London , and hence complying with the rules of standard British English instead of General American pronunciation as far as rhoticity is concerned.
10. However, nonrhoticity in NAE stayed rather variable – in contrast to the standardized nonrhoticity of British English – and persisted as a prestige norm in NYCE, used rather arbitrarily, as Hubbell formulated in his statement.
11. Thus, the use of r in final and preconsonantal positions can be seen as a paragon of free variation , which, however, is highly influenced by sociolinguistic features and social variables.
12. These social stratifications have been sufficiently studied in various experiments and are to be presented hereafter.
13. In 1962, Labov triggered sociolinguistic studies of the pronunciation of r in NYC by conducting his New York department store study, analyzing social, ethnic and gender influences on the speech of New York City inhabitants.
14. The experiment was carried out in three department stores of varying social status, Saks, Macy’s and Klein , where Labov approached the informant in the role of a customer asking for directions to a particular department.
15. The target answer for assessing rhoticity was Fourth floor , as both words feature final and postvocalic rs, which are typically omitted in traditional New York City speech.
16. Labov , however, regarded various variables, which could possibly lead to a use of a rhotic variety of the utterance: sex, age , occupation , race, foreign or regional accent, if any.
17. This method and the analysis of all 264 interviewees , which will be presented hereafter, still functions as one of the most important assessments of social stratification of r and has been reproduced several times.
18. The most recent replica has been conducted by Patrick-André Mather in 2009, providing a possibility to analyze the diachronic development of the use of r in postvocalic position between 1962 and 2009 .
19. Moreover, Labov conducted a study concerned with rhoticity in the Lower East Side of Manhattan, as this particular area is characterized by its various ethnic and socioeconomic groups , broadening the possibilities of social analysis of the pronunciation of r .
20. Becker revisits this particular field site, in order to study the diachronic change in NYCE as well.
21. Generally, taking all the results of studies into examination, it is obvious that rhoticity is spreading in the variety of NYCE.
22. Especially Mather proved an increase in the usage of r-1, especially among young speakers of NYCE , whereas traditional, solely r-0 utterances, which were primarily produced by lower class inhabitants of NYC and in spoken language among other social groups, seem to decrease rather rapidly, as Becker demonstrates.
23. This change to rhoticity can be attributed to various factors and developments, however, this paper highlights the variation of r due to ethnic influences and ethnolectal changes.
24. New York City English is influenced by many ethnic varieties, shaping the city’s dialect to its diverse form.
25. Thus, nonrhoticity and the recent change to rhoticity in NYC can also be described through ethnic factors, adding up to the already consisting social reasons.
26. In Labov’s studies , he found three ethnic groups among his interviewees to be discussed apart from white residents, namely Jews, Italians and African Americans .
27. However, Jews and Italians were considered to follow the same rules as far as rhoticity was concerned as white New Yorkers and other European immigrants , whereas the question whether ethnicity does influence rhoticity remains open.
28. In Labov’s Department Store Studies , but also during his field work in the Lower East Side , however, certain patterns were found considering African American speakers of NYCE, as they were found to produce lower rates of r-1 than their white counterparts , leading to assumptions, that African American Vernacular English rather features nonrhoticity than rhoticity.
29. Especially the northern region of Manhattan around the neighborhood of Harlem is highly influenced by African American inhabitants and their dialect, however, free variation of nonrhoticity in AAVE is nowadays rather acknowledged as an ethnolect rather than a feature of NYCE dialect .
30. Still, African American speakers are considered to perform the change to rhoticity as much as white inhabitants .

31. However, Mather’s replica in 2009 also proved that African Americans do not use rhoticity as much as white interviewees , which he stresses.
32. Thus, Mather summarizes the output of his replica, emphasizing the role of AAVE in terms of nonrhoticity and African Americans’ role in maintaining a certain nonrhotic variety in New York City.
33. Allophones as variations of phonemes are usually sensitive only to the phonetic environment .
34. However, there are cases, where allophones are not entirely bound by phonemic rules, but rather by morphological conventions or even by word classes as well , complicating the correct use of allophones for non-native speakers of a language, as such variations are almost impossible to acquire in a 2nd language.
35. In German, for instance, the variation of allophones of the so-called Umlaut is influenced by morphology , aggravating the study of German for non-native speakers.
36. Such a phenomenon is also common in New York City English, as its speakers have a tendency to produce a tense allophone of ae before voiceless fricatives, voiced stops and front nasals .
37. Especially in words such as cab, bad, or gag , this higher allophone of the phoneme ae can be identified as traditional New York City English, especially associated with a Brooklynese dialect , however, this characteristic is not limited to a certain neighborhood, but rather constitutes traditional New York City pronunciation .
38. To investigate, whether the short a system is still current in New York City English, Kara Becker and Amy Wing-mei Wong conducted a study measuring each of 2355 speakers acoustically and analyzed their pronunciation of ae to assess whether the traditional short a was still popular in NYCE.
39. As a first distinguishing method, interviewees were sorted into three categories according to age , and as a further measure, the largest ethnic minority groups in New York City: African Americans , Chinese, and Puerto Rican were taken into account in order to be able to acquire a representative sample of New York City inhabitants.
40. After having conducted the study, it became obvious for Becker and Wing-mei Wong that older native speakers of NYCE are more likely to produce a traditional short a, tensing the vowel before voiceless fricatives, voiceless stops and front nasals .
41. Based on age groups, the frequency of articulating the classic New York ae decreased from then downwards, as younger native New Yorkers tend to produce a tense ae only before nasals, whereas in other position the classic NYCE ae is hardly or not at all audible .
42. As discussed previously, ethnic groups do influence language itself and especially dialects and so it is in New York City’s pronunciation of ae.
43. Becker and Wing-mei Wong found strictly speaking 4 ethnic groups to be discussed as far as pronunciation of ae is concerned, namely white New Yorkers, African Americans, Chinese and Puerto Ricans .
44. In Labov’s Lower East Side studies from 1962, in which all peculiarities of New York City English were examined, however, there was a distinction made between whites, African American and European groups, such as Italians, as well as religious minorities such as Jewish or Orthodox inhabitants .
45. Nonetheless, as a common thread, both studies made obvious that the city’s ethnic groups do influence and shape New York City English- but only to a certain extent.
46. The contemporary research by Becker and Wing-mei Wong proves that all three ethnic groups examined do not differ notably from the results of white New Yorkers of all age groups, as, for instance, younger New York citizens with whichever ethnic background do not articulate the classic NYCE short a .
47. Observing Labov’s classic Lower East Side study, however, it is made obvious that there are differences between ethnic groups pronouncing aeh , especially when it comes to a comparison between Italian and Jewish citizens of New York.
48. While Italian speakers of NYCE statistically use aeh more often, Jewish citizens are very unlikely to produce the classic New York ae .
49. Since the Labov study took place in the 1960s, and Becker and Wing-mei Wong ensured in 2010 that all interviewees were native speakers of English , recent studies presumably tend to be more valuable when it comes to investigating ethnic influences on the pronunciation of ae, as the city’s population has naturally changed over the intervening years as well.
50. Still, all studies concerned with this topic conducted in New York City are constructive measures to evaluate and reevaluate the connections between ethnicities and the New York City English’s short a .
51. A city as big and diverse as New York contains a vast amount of dialects, ethnolects and linguistic phenomena.
52. New York City English as such, therefore, is constituted by all those factors, making it impossible to speak nowadays of only one unified dialect of North American English.
53. After having studied the most striking features attributed to the traditional NYC speech, it became quite obvious that such a cohesive NYC dialect, which existed for most parts of the city’s history, is on the decline, as both, the nonrhotic variety of r and the tense ae are rather found in the older population stratum.
54. Younger native speakers of NYCE, therefore, tend to produce General American pronunciation, with only glimpses of traditional NYCE features.
55. Also, regarding ethnicity, minority groups were not found to influence the city’s dialect significantly, as the overall trend in NYCE is borne by all ethnic groups to equal portions.
56. Only African American Vernacular English was found to influence the persistence of nonrhoticity in New York City English.
57. Though, on the whole, native speakers of NYCE with whichever ethnic background do not bring certain features into the dialect, but generally comply with the common tendencies.

Sample 28, SE paper text

1. Through the course of history, the political sphere of society has used language as a tool to people’s hearts and minds.
2. As a result, language has been used and misused in order to seek power, which can be seen in the use of language in the field of propaganda.
3. In the following paper, I will not only define political discourse and political language in general, but also describe the language of propaganda from a linguist’s point of view as well as discuss theoretical models and practical examples of propaganda usage.
4. When analyzing the language of propaganda, one has to define the meaning of the latter in the first place.
5. The Cambridge Dictionary, for instance, explains the term as information, ideas, opinions, or images, often only giving one part of an argument, that are broadcast, published, or in some other way spread with the intention of influencing people’s opinions .
6. Therefore, this definition encompasses all aspects of human communication, from advertising to politics, and could rather be described as a general notion or broad definition.
7. Even though the term propaganda could also be used for broadcasting information in order to manipulate subjects to purchase items , a more narrow definition that suits the purpose of this paper better could be found in the Oxford Dictionary.

8. Here, propaganda is defined as information, especially of a biased or misleading nature, used to promote a political cause or point of view .
9. It can be seen that the political function of propaganda is foregrounded, which not only emphasizes this paper's purpose, but also the main function of propaganda in society.
10. Therefore, propaganda could also be described as ways of using language in political discourse , under the premise of Oxford Dictionary's assumption of bias and misdirection.
11. As a result, however, one could say that political propaganda is rather a branch of political discourse, which will be discussed in Chapter 3.
12. Political discourse primarily focuses on the meaning which goes beyond what is being said .
13. Therefore, political language and discourse are very often analyzed in the field of pragmatics, in which the implied meaning is the center of analysis .
14. As a result, this field of study is very often – particularly by more traditional linguists who emphasize the importance of form over function – described as vague or confusing .
15. However, pragmatic linguists are typically studying meanings within one particular context , which, therefore, narrows down its scope as well as reduces possible vagueness of results.
16. In this particular case, the study of meaning concentrates on political language and its primary function of sharing political ideas and ideologies .
17. One sub-field that is concerned with these functions of political language is the field of critical linguistics, in which the relationship between language and reality is foregrounded .
18. As a result, critical linguists consider the role of language in the creation and maintenance of political and social ideologies .
19. This definition, consequently, leads to the supposition that the view on the world as well as on society may vary according to the ideology that is presenting it.
20. Therefore, language can not only be seen as a tool to express and share a certain political idea, but it also gives room for different interpretations of the world and everything in it .
21. As a result, critical linguists proceed on the assumption that there is no one all- encompassing reality which is adjusted to the respective ideology, but rather several opposing realities, on which various political and social ideologies are built upon .
22. Therefore, different, and in most cases contradictory, realities are offered to society.
23. In order to share and spread these oppositional ideas, political parties need to make use of language as a means of expression and persuasion.
24. To understand how language functions in a political discourse, however, one has to grasp the overall possibilities of language as a means of communication in the first place.
25. Language can not only been seen as a strict system of syntactic structures, but as a means of communication .
26. As a result, meaning is messaged through grammatical forms in order to express thoughts, opinions, ideas and suggestions .
27. Consequently, we as users of language are able to discuss, exchange, debate, agree, disagree and come to conclusions by using persuasive and convincing language items to win over another person.
28. It is without a doubt in the nature of the human race to not only express oneself, but also to try to convince the opposite in order to be right.
29. This innocent concept of discussion and exchange, however, can also be misused in order to manipulate and influence people's free will for personal enrichments such as power.
30. At the same time, language as an expressive tool also offers the possibility of spreading falsehoods or certain ideologies for the already mentioned pursuit of control and authority over society.
31. This misuse of the communicative and expressive functions of language is, however, not a modern concept.
32. As Wilson stated, it is since classical times that it has been accepted that language plays a role in the creation of political reality .
33. Therefore, one could argue that humanity simply accepts the fact that political leaders use these persuasive elements in language in order to convince citizens of certain ideologies or parties.
34. Consequently, this norm or standard, which can be found across all borders and continents, gives room for possible misuse of persuasive elements as well.
35. This concept has been taken up by several linguists, who state that political messages are not only explicit, but also carry implicit meaning through which individual listeners may be influenced or even manipulated .
36. Typical features of political language have been defined by the famous author of Nineteen Eighty-Four and Animal Farm George Orwell.
37. In his essay Politics and the English Language , Orwell stated that political language – and with variations this is true of all political parties, from Conservatives to Anarchists - is designed to make lies sound 5truthful and murder respectable, and to give appearance of solidity to pure wind .
38. Even though this description was used for the English political discourse, it certainly can be adapted to any language on earth.
39. Its main characteristics have further been described by Orwell.
40. Political language particularly reveals its characteristics as well as its raison d'être in the most pressing issues of society.
41. As one of the most influential topics, especially in today's globalized society, ethnic affairs are likely to provide insight in common practices of political discourse.
42. In a democratic state, participants of political discourse are not only governments and parliaments, but all political parties, bureaucracies as well as other political organizations who influence political progress and discussion .
43. Thus, through the use of political language, these subjects control prominent topics in ethnic relations such as Increasing illegal immigration, waves of refugees, housing, ghettos, black crime, minority unemployment, Affirmative Action, multicultural education, poverty and disadvantaged minority groups, as well as discrimination and racial hate directed against minorities by white majority groups.
44. Even though mainstream political leaders – in contrast to right-wing extremists – typically approach these topics with political language characteristics, such as vagueness or euphemisms, they can still be considered as part of the problem .
45. Therefore, one could argue that the excessive use of political language not only contributes to the manipulation of individuals in order to gain support, but also to oblivion as far as ethnic issues of a society are concerned.
46. At the same time, politicians are eager to deny their role in the spread of racism and discrimination.
47. Especially through the use of glorious self-descriptions, the underlying tolerance of a society is emphasized through which all sorts of discrimination should be proved as impossible .
48. However, these statements are often followed by a more negative portrayal of minorities or even other nations, which require action with negative effects for these very subjects .

49. Therefore, political language can turn on specific classes of population or other cultures, but never against the majority of a society, who is voting in favor of the party in charge.
50. Van Dijk appropriately explained this phenomenon in the following way: The general structure of disclaimers in discourse about ethnic affairs usually is We are positive, but They are negative, as in We are very tolerant, but They abuse our tolerance.
51. Or it may start with a denial of a negative property of the own group followed by a negative property attributed to others, or followed by a negative decision, as in, We have nothing against immigrants, but we can’t let everybody in.
52. Even though this statement was written in 1993, one can see that the general notion of ethnic affairs in political language is still the same – and more relevant than ever.
53. Especially negative other-presentations dominate the contemporary political discourse across borders in Europe and in the United States, through which racist bodies of thought can grow and discrimination against people of color is on the rise yet again.
54. The political institutions have yet to come up with a more inclusive political language in order to prevent nationalist ideologies to dominate the political landscape.
55. A traditional approach towards political language however, does not seem to be the best formula regarding societies’ most pressing issues.
56. Political success, however, is also highly dependent on the politician who is seeking power.
57. Especially in times of crisis, it can be argued that a person who embodies national values is likely to be regarded as charismatic by the public , hence successful.
58. This theory can be seen in practice in the US-election of 2008 as Barack Obama personified the classic American Dream in the face of a global economic crisis.
59. Obviously, charismatic leadership is to be contrasted with authoritarian systems in which the public has no say in the decision .
60. As a result, political language used by charismatic leaders of free and democratic nations can be clearly distinguished from totalitarian language .
61. The world-famous persona of Abraham Lincoln can be made an example when it comes to charisma in political discourse.
62. Even though his speeches and the myth around his person came to fame up until now, contemporaries would not have described him as a particularly charismatic person .
63. Many portrayed him as awkward in appearance and plain with a rather unanimated voice .
64. However, his speeches were regarded as exceptionally clear and full of logical and intellectual arguments no one has ever thought of before .
65. During times of deep troubles, his speeches calmed the nation down, and, therefore, embodied not only charisma, but also the American values that their differences were not irreparable and that both sides must come together .
66. It is questionable whether Lincoln would have gotten the status of a hero if he had been forced to speak on national television, but still, his story reveals the importance of charisma in political language of democratic nations.
67. A similar example, yet with an entirely different approach towards charisma, can be found when analyzing the political career of former President of Egypt Gamal ‘Abd al-Nasser.
68. As a result of the modern times, he identified the people of Egypt as a heterogeneous group consisting of different social as well as religious spheres .
69. Therefore, he used different forms and degrees of political language according to the target audience he was speaking to .
70. Furthermore, he broke the tradition of using classical Arab and adjusted his dialect to the respective region or social class .
71. As a result, his ever-changing register of political language not only led to the creation of a multifaceted persona in charge, but also guaranteed Nasser a charismatic legacy he still occupies in Egypt up to his day.
72. Regarding charisma, it is impossible to leave out former President of South Africa and Nobel Peace Prize Laureate Rolihlahla Nelson Mandela .
73. As an activist for democracy, equality and peace he used his extraordinary rhetoric talent to gain attention from South African citizens as well as the entire world.
74. Elleke Boehmer defined this special talent as an ability to inspire undivided attention and devotion in whomsoever he was speaking to, something that his fans and followers soon came to call Madiba Magic.

Sample 57, PS paper text

1. Every person on this planet communicates, but how is that possible?
2. The question of how humans acquire language is of high interest to linguists and and has been the topic of much research.
3. The purpose of this paper is to look at the process involved in language acquisition and will focus on nativists´ explanation of the learning process and defense of their theory.
4. In order to do this, section 2 will provide insight on how nativists like Chomsky define and approach language acquisition.
5. Further along, section 3 will illustrate the two main arguments used to support this approach, namely the poverty of stimulus argument, and the critical period hypothesis.
6. While reviewing the critical period hypothesis, the section 3.
7. 1 will focus on two famous examples of testing this argument, namely the Genie case as well as testing deaf babies.
8. In section 4 a conclusion will be presented to close the paper.
9. It will become apparent that while both the poverty of stimulus and the critical period hypothesis are useful to some extent to describe the nativist approach, neither is sufficient enough to actually provide evidence or explain how exactly the language acquisition process works.
10. This will lead to a call for more empirical research.
11. Before reviewing any arguments that try to demonstrate the innateness theory, ‘language acquisition’ itself needs to be defined.
12. There are a variety of experts on the field that have explained language acquisition, however the problem is that how they think language acquisition works is linked to the theory they are trying to convey.
13. Since the focus of this paper is on the nativist approach, the plausible action is to look at how nativists have defined it.
14. Chomsky, for instance, explains language acquisition as a process that simply happens to children without their knowledge, not unlike any other processes that humans undergo without being aware of it.
15. He compares this procedure to the sexual maturation, another progression that children do not actively think about or activate, but rather one that simply happens when the time and environment is appropriate .
16. Yang explains the naturalistic approach by stating that language occurs due to our brain, which is an organ that grows and develops naturally, or biologically.
17. Therefore, language and how humans learn it has to be studied according to natural science´s guidelines .
18. All of these definitions beg the question how this could possibly work and what makes linguists support the nativism approach.

19. Among the arguments typically drawn on in support for nativist theories of language acquisition, the most prominent ones are the poverty of stimulus argument and the critical period hypothesis.
20. In the following, both of these theories will be discussed in more detail, assessing the usefulness in explaining and supporting the nativism approach.
21. To put it simply, the POS refers to the relationship between input to language learners and their linguistic competence .
22. The term ‘the poverty of stimulus’ was brought to the forefront by Chomsky in the 1950s, however, back then the concept was not properly developed yet, as it did not challenge the appropriateness of the input .
23. From the 1970s onwards, a wave of generativists started researching and found examples for the notion that children have and produce grammar that has not been explained to them .
24. Linguists like Peter and Baker have proceeded to elaborate the POS argument and introduced the new term ‘projection problem’ .
25. The projection problem revolves around the question what the functional relation that exists between an arbitrary human being ´s early linguistic experience and his resulting adult intuition is .
26. Baker outlines two main methods to explain this.
27. The first one is that this intuition can be explained using relatively few principles that are assumed to also be part of numerous learning processes apart from language, for instance analogy or similarity.
28. The other approach concludes that linguistic intuition arises from an inherent disposition, commonly referred to as ‘universal grammar’ in the generative literature.
29. Over two decades of research later, Pullum and Scholz sum up a very strong POS theory from Lightfoot, which is that children learn some aspects of language without any access to it.
30. This means that despite never hearing some rare language features children are able to produce them.
31. This implies that children have to have a non-experience based language acquisition process .
32. However, Pullum and Scholz also point out that this definition is circular, since it explains the native theory by presupposing nativism.
33. This is an issue about the poverty of stimulus argument that, before Pullum and Scholz, has not been addressed by any nativist linguist , which leads to the conclusion that even after half a century of researching the poverty of stimulus argument, there still are questions about how it works and why, and even if the argument actually holds.
34. Since the poverty of stimulus argument has not been able to properly explain the innateness theory, a different approach is needed.
35. The second influential concept that is used to defend nativism is the critical period hypothesis.
36. Saxton suggests that the critical period can be defined as a span of time during which humans are plastic in their development, meaning there are parts of a humans´ life during which they will react quickly and more efficiently to input.
37. Since the development of humans has an end point, and that end point is rather early in people´s lives, critical periods are set to be early on as well .
38. Saxton addresses the issue one encounters when trying to test the critical period hypothesis, namely how highly immoral it would be to deprive infants of any linguistic contact whatsoever in order to test whether or not they would acquire a language by themselves or if a delayed learning process would have any serious consequences .
39. Supposedly, there are four instances where such an experiment took place before the 18th century, however, only one of those took place in a scientific and academic surrounding, which begs the question whether the others are results that can be taken into account when dealing with the critical period hypothesis, or merely ancient stories .
40. There is, however, one highly referenced example of a child living in linguistic isolation: Genie.
41. This girl, according to Saxton , was held captive in a bare room by her parents who only rarely spoke to or in front of her, and every time she tried to vocalize on her own her father would assault her.
42. She was found in the nineteen seventies when she was thirteen and at that time was unable to speak at all; therefore, linguistic interest in her was high.
43. She was the ideal example to test the critical period theory on.
44. It was found that the critical period is more or less effective in different aspects of language .
45. Saxton was able summarize observations made by the researcher working with Genie, and found that with regards to phonology, there seems to be evidence that this is part of its own critical period, especially considering that any L2 learners struggle or are unable to acquire a native-sounding accent after they are five years old, and Genie was no exception .
46. Looking at lexicon it was noticeable that Genie did in fact learn several hundred words, which would disprove the critical period hypothesis; however, this seems very insignificant in comparison to any ordinary two-year-old, who not only learns more words but also learns them more quickly .
47. This prompts the assumption that if vocabulary learning has its own critical period, it may be considerably longer than the one for phonology.
48. Concerning morphology, it was found that Genie seemed to be lacking inflectional morphemes, which led to her uttering sentences like ‘small two cups’ and ‘like chew meat’ .
49. Thus, it is apparent that morphology severely suffers when it is acquired after a certain point of age.
50. The last issue Saxton points out is Genie´s acquisition speed.
51. Usually children enter the two-word phase after having acquired 50 or so words and rest there for approximately seven weeks.
52. Genie, on the other hand, learned 200 words before she started using two-word phrases and then she spent five months in this stage, almost three times as long .
53. This may seem like certain proof that there is a critical period, because this girl was not in contact with language during the early stages or sensitive periods and the result is that she had severe difficulties with language learning in various aspects and struggled significantly more with the language acquisition process than toddlers do.
54. However, it must not be forgotten that she lived in isolation, and therefore it is challenging if not impossible to detect whether or not her difficulties actually resulted from the lack of linguistic input, since her entire development was impaired due to the inhumane way she was raised .
55. Since it is considered cruel and unethical to deliberately rob children of their linguistic input, researchers have tried to find other ways of testing the critical period hypothesis.
56. Cormier et al state that there are only two methods of analyzing the critical period hypothesis: one is restricting children’s access to linguistic input, and the other is testing babies that have been born deaf .
57. The first approach has been dealt with by referring to Genie, and the second one will be presented now.
58. Purves and his colleagues found that the language development of deaf babies who have not been exposed to any nonverbal language suffer severely, whereas congenitally deaf babies that have been exposed to some sort of sign language will start to communicate and babble with their hands .

59. This implies that there is a critical period during which children learn language faster and more efficiently, as well as that kids need some sort of linguistic input.
60. This could be an indicator that language develops and if something hinders that development there are consequences, similar to the growing process of a human that happens naturally and without our awareness, which could be used as a solid argument for the nativists´ approach to language acquisition.
61. Taking into account the fact that the poverty of the stimulus argument did not succeed in justifying the nativism theory, but the critical period argument definitely provides testified results in favor of nativism, it can be concluded that there certainly are aspects of the innateness approach that seem plausible.
62. The problem so far is that the critical period hypothesis is only descriptive and the poverty of stimulus argument, which tries to explain why the nativism approach is the only correct one, ends up being incomplete or circular.
63. Therefore, a lot more research is still needed if one is looking to once and for all settle the ongoing discrepancy between nativists and non-nativists.
64. This paper has found that it is quite difficult to define language acquisition, because the definitions used by different authors will always be linked to the approach the writer is defending.
65. Further, it was discovered that the nativist approach is concerned with the natural and automatic acquisition of language, which means that representatives of this theory believe that language is something that develops with the body and from the body.
66. The two main arguments for the nativist theory have been brought up.
67. The first one has been the poverty of stimulus argument, which implies that children learn language and language structures without or with limited input.
68. The second one has been the critical period hypothesis, which states that there is a part of children’s lives during which they are especially sensitive to language learning.
69. In order to prove this, this text explored the infamous Genie case, as well as the research done on deaf babies.
70. Unfortunately, only a handful of aspects could be looked into and these insights were only partly satisfactory in actively proving the innateness approach.
71. It was not possible to deduce from the poverty of stimulus argument that the approach of nativists is in fact true, and the critical period theory only led to the assumption that there are some aspects of language learning that are time sensitive, but did not explain why or how.

Sample 57, SE paper text

1. Language Aptitude is a research topic, which is studied by several fields.
2. Due to the practicality of language aptitude testing, this area has been extensively explored and funded.
3. However, there are still areas that need further investigation.
4. In this paper, the correlation between language aptitude and language learning styles will be researched.
5. To do this, the hypotheses people with a reflective, sensing, visual or sequential learning style will have better LLAMA B test scores and people with a more developed learning style profile will have better LLAMA B test scores will be tested in order to answer the overall research question, which is how do the LLAMA B test and language learning styles correlate.
6. Before this can be achieved, however, it is necessary to establish what language aptitude is, what learning styles are and how they could be tested.
7. As Wen summarizes, research into foreign language aptitude is now approximately 70 years old, and one of its main accomplishments is the development of the Modern Language Aptitude Test .
8. The Modern Language Aptitude Test is a test designed by Carroll and Sapon and will be further described in the section Testing Foreign Language Aptitude.
9. In Carroll’s research of language aptitude , he distinguishes four main elements that amount to language aptitude, namely phonetic coding, grammatical sensitivity, rote memory for foreign language materials, and inductive language learning ability.
10. The most important component out of these is phonetic coding, which can be explained as people coding auditory material in a way that it can be recognized or reproduced after a while.
11. This ability is essential for language learning, as people that lack it would have difficulties with both remembering and memorizing phonetic elements.
12. Grammatical sensitivity describes how people maneuver grammar, the function and grouping of words and sentences.
13. Rote memory for foreign language materials can be defined as the ability to learn association pairs, and inductive language learning ability is the ability to infer linguistic forms, rules, and patterns from new linguistic content itself with a minimum of supervision or guidance .
14. The most important way for language aptitude research of explaining what the key features that comprise foreign language aptitude of has been presented, but the general question what language aptitude is has yet to be answered.
15. Abrahamsson and Hyltenstam define it as a largely innate, relatively fixed talent for learning languages that is rather independent of general intelligence, musical talent, motivation to learn a language and attitude towards language learning.
16. Wen, Biedron and Skehan summarize previous definitions into foreign language aptitude being an umbrella term consisting of a set of cognitive abilities, thus making it a componential concept .
17. Aptitude therefore is a combination of different cognitive abilities that help language learning, rather than an individual talent .
18. According to these definitions, aptitude can be used synonymously with ability .
19. It is imperative to note that language aptitude research is a multi-field research topic, with the most significant ones being applied linguistics, educational psychology, cognitive psychology, and cognitive neuroscience .
20. The field with the strongest influence is the education sector, as language aptitude is applied here.
21. Despite the immense changes that the educational practice has gone through in the time that language aptitude has been researched, it still remains a valid concept today .
22. However, interest in the research only increased again at the beginning of the century.
23. Before that turning point research stagnated at Carroll’s principle of language aptitude as a somewhat reliable prediction of the likely learning rate and outcomes .
24. Now there are several new approaches to language aptitude, some of which will be presented in the following section.

1. As mentioned above, the Modern Language Aptitude Test was established by Carroll and Sapon in 1959.
2. The MLAT has been used for selection, placement, and guidance by schools and US government agencies for more than 40 years .
3. It was developed due to the army, and consequently the government, requiring a test that would be able to predict language learning abilities, as they needed to hire people that would be able to learn foreign languages quickly .

4. As for how the MLAT is designed, Sasaki explains that the MLAT is aimed at English-speaking adults and is part of the language aptitude test family, which also includes a British version of the test and the MLAT-Elementary for younger people.
5. Due to copyright, researchers that chose to use the MLAT have to be given permission first.
6. The test is divided into 5 parts which assess at least one of Carroll's four main components of language aptitude that are explained in the section above.
7. In the first part, participants are presented orally with numbers in an unknown language.
8. Then they have to translate new number combinations into English.
9. The second part is a sound-symbol correlation test aimed at testing phonetic coding ability.
10. The third part includes the participants receiving spelling clues that they have to semantically match to a word.
11. The fourth part examines grammatical abilities, as the participants have to find a word in a sentence with the same function as the one they are presented with.
12. The last part consists of pair associations that have to be memorized.
13. The completion of the entire test takes about 70 minutes, however it is possible to only do part 3 to 5.
14. The LLAMA Language Aptitude Test was created by Meara in 2005 and is based on the MLAT .
15. The main advantage of the LLAMA test is that it is not language dependent, unlike the MLAT, as it uses a fictitious language and visual input .
16. It is a program that is free to download and can therefore be used by any researcher without the necessity of being granted permission first.
17. It consists of four parts, namely LLAMA B, LLAMA D, LLAMA E, and LLAMA F .
18. The LLAMA B focuses on vocabulary and is explained in more detailed on the section Instruments.
19. In the LLAMA D, the participants are presented with foreign sounds that they have to recognize later on .
20. The LLAMA E is a sound-symbol association test.
21. The participants are presented with sounds and symbols of a different alphabet and are tasked to remember association pairs .
22. The last test, LLAMA F, measures the ability to infer or induce the rules of an unknown language .
23. The Cognitive Ability for Novelty in Acquisition of Language Test requires participants to learn aspects of an artificial language .
24. Thus, the test investigates the learning process as it is happening, with the subjects starting without any knowledge of the fictitious language and ending with the ability to comprehend a short story .
25. As Biedron and Pawlak summarize, this test is based on the conceptualization of language aptitude as a dynamic construct, which interacts with the environment, depends on evolutionary development and is potentially trainable .
26. As such, the test allows for different learning mechanisms, for instance explicit or implicit learning, and explores learning strengths and weaknesses .
27. The test itself consists of nine sections, five of which are completed immediately, and four at a later date .
28. The sections are as follows: In section one participants had to learn meaning from context, in section two the task was to understand the meaning of a passage, in section three the subjects had to learn association pairs, and in section four the goal was sentential inference.
29. All of these sections were completed once and then repeated at a delayed time.
30. Lastly, section five consists of learning language rules and was only presented once .
31. Similarly to the MLAT, the High Level Language Aptitude Battery arose out of military despair .
32. After the nine-eleven attacks the American military was in desperate need of more language experts, which resulted in more funding for research and a new language aptitude test .
33. The difference to other language aptitude tests is that it is targeted at talented, high-level learners .
34. Due to this, the empirically observed phenomena of people reaching their personal limits at a certain proficiency even with high aptitude scores can be overcome .
35. The test is rooted in contemporary cognitive science and examines mainly the working memory, but also the short-term memory and the longer-term memory .
36. It was initially designed to test the short-term memory capacity, executive capacity and control, rote memory, perceptual acuity, processing speed, priming, implicit induction, explicit induction, pragmatic sensitivity and fluidity and was then shortened to examine the working memory, associative memory, priming, implicit learning, processing speed, and auditory perceptual acuity .
37. Dörnyei defines learning style as a concept that represents a profile of the individual's approach to learning, a blueprint of the habitual or preferred way the individual perceives, interacts with and responds to the learning environment .
38. He further explains that learning styles have soft boundaries and are therefore not easy to classify.
39. They are independent of situational influence, and yet they appear to be influenced by situation to some degree.
40. Further, it is difficult to distinguish what constitutes personality and learning styles, which leads to the assumption that learning styles are one part of what defines personality .
41. As the definition of learning styles can only state that it is about characteristic strengths and preferences in the ways people take in and process information , there are several different models that test learning styles.
42. As Felder summarizes, this model classifies participants based on the task-specialized functioning of the physical brain.
43. There are four quadrants, namely quadrant A to D .
44. Quadrant A targets the cerebral left brain and is used for logic and analyzing, as well as quantitative, factual and critical thinking.
45. Quadrant B targets the limbic left brain, which is applicable to sequential, organized, planned, detailed, and structured thinking.
46. Quadrant C aims for the limbic right brain and regulates emotional, interpersonal, sensory, kinesthetic, and symbolic processes.
47. Lastly, quadrant D is the cerebral right brain, which is responsible for visual, holistic, and innovative activities .
48. As Ayoubi and Ustwani explain, the Myers-Briggs Type Indicator tests on four levels: source of energy, processes of perception, processes of judging, and attitudes towards the outside world.
49. This results in eight preferences, namely introversion or extroversion regarding the source of energy, sensing or intuition regarding the process of perception, thinking or feeling regarding the process of judging, and perceiving or judging regarding attitudes towards the outside world.
50. Consequently, there are 16 personality types possible.
51. Kolb's learning style model is based on how people process and integrate information, and on non-dominant modes of expression .
52. As Jones, Reichard and Mokhtari further explain, Kolb combines a horizontal axis of perceiving with a vertical axis of processing, and by placing the axes within a circle he created four distinct learning modes that represent different types of learning: concrete experience, reflective observation, abstract conceptualization, and active experimentation .

53. If a person combines concrete experience with reflective observation they are divergers, a combination of reflective observation and abstract conceptualization is most useful to assimilators, abstract conceptualization and active experimentation are practical for convergers, and active experimentation and concrete experience is best for accommodators .
54. This version classifies people into 5 continuous scales, namely sensing learner and intuitive learners, visual and verbal learners, inductive and deductive learners, active and reflective learners, and sequential and global learners .
55. Sensing learners are practical and prefer working with facts, whereas intuitive learners are innovative, conceptual and prefer theories and concepts to single facts.
56. Visual learners favor visual presentation of material, while verbal learners prefer written or spoken explanation.
57. Inductive learners work better when information is presented starting with the specific and proceeding to the general, whereas deductive learners favor a process from the general to the specific.
58. Active learners thrive when working with others and trying things out, and reflective learners learn by thinking things through and working alone.
59. Lastly, sequential learners learn in small linear steps, whereas global learners learn in large chunks .
60. The test that is based on this theory lacks the inductive and deductive learners, and will be further explained in the section Instruments.
61. The study was conducted using two online tests.
62. Therefore, the participants were allowed to complete these tests in the comfortable atmosphere of their home and without disturbances or time pressure.

6 Parse tree example used for comparison

The parse trees for each sentence have been compared with the manual annotation and according to the definitions as described for each of the defined sequencing units as Tregex patterns in Lu (2010: 480–483) to explain possible differences in the annotations. This helped to raise the validity of the L2SCA and the manual annotations. In the following tables the annotations for each sentence of the compared samples can be found and the comments point out parsing problems and other problems in detecting the correct sequencing units.

Parse tree example from Stanford parser for sample 1, SE paper, sentence 51:

Herdina and Jessner argue that monolinguals, bilinguals and multilinguals employ different strategies when learning a foreign language.

```
(ROOT                                             # marks new sentence & T-unit  → S=1, T=1
 (S                                               # new clause
  (NP (NNP Herdina)
   (CC and)                                       # marks coordinated phrase     → CP=1
   (NNP Jessner))
  (VP (VBP argue)                                 # clause headed by tensed verb →VP=1, C=1
   (SBAR (IN that)                                # marks dependent clause
    (S                                            # new clause
     (NP (NNS monolinguals) (, ,) (NNS bilinguals)
      (CC and)                                    # marks coordinated phrase     → CP=2
      (NNS multilinguals))
     (VP (VBP employ)                             #clause headed by tensed verb  →VP=2,C=2,DC=1
      (NP (JJ different) (NNS strategies))
      (SBAR                                       # marks dependent clause
       (WHADVP (WRB when))
       (S                                         # new clause
        (VP (VBG learning)                        # no tensed verb               →VP=3
         (NP (DT a) (JJ foreign) (NN language)))))))))
```

Results of the L2SCA for sentence 51 of SE paper from sample 1:

Words (W)	Sentences (S)	T-units (T)	Verb phrases (VP)	Clauses (C)	Dependent clauses (DC)	Coordinated phrases (CP)
17	1	1	3	2	1	2

Abbreviations of the Stanford parser:

VP = verb phrase
VB = verb base form
VBG = gerund or past participle
VBD = past tense
VBN = past participle
VBP = non-3rd person singular present
VBZ = 3rd person singular present
NP = noun phrase
NN = noun singular or mass
NNS = noun plural
NNP = proper noun singular
JJ = adjective
CC = coordinating conjunction

7 Manual and L2SCA annotations

The following pages show the manual and the L2SCA annotations for each sentence in the annotated texts. Differences in annotations are marked with yellow shading. See the parse tree example in the above part of the appendix for an explanation of the abbreviations used in the comments.

Table 10: Comparison of annotations for PS paper of sample 1

Comparison Sample 1 / PS text (page 1/2)															
	Manual_annotation							**L2SCA_annotation**							**Comments**
Sent.	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	
1	26	1	3	2	1	1	1	26	1	3	2	1	1	1	
2	19	1	3	3	3	0	0	19	1	3	3	3	0	0	
3	22	1	2	2	1	1	1	22	1	2	2	1	1	1	
4	18	1	2	2	2	0	0	18	1	2	2	2	0	0	
5	16	1	1	1	1	0	0	16	1	1	1	1	0	0	
6	40	1	3	2	1	1	2	40	1	3	1	1	0	2	VBP tagged as VB → DC not dedected
7	9	1	1	1	1	0	0	9	1	2	1	1	0	0	“to” tagged as VP
8	23	1	3	2	1	1	0	23	1	3	2	1	1	0	
9	17	1	1	1	1	0	0	17	1	1	1	1	0	0	
10	36	1	6	4	3	1	0	36	1	6	4	3	1	0	
11	25	1	2	2	1	1	0	25	1	2	2	1	1	0	
12	36	1	4	4	1	3	1	36	1	4	4	1	3	1	
13	20	1	2	2	1	1	0	20	1	2	2	1	1	0	
14	37	1	3	3	1	2	2	37	1	4	3	1	1	2	latinate word/ “approaches” tagged as VP
15	35	1	3	3	1	2	0	35	1	3	3	1	2	0	
16	20	1	2	2	1	1	0	20	1	2	1	1	0	0	VBP tagged as VB → DC not dedected
17	19	1	2	2	2	0	1	19	1	2	2	2	0	1	
18	18	1	2	1	1	0	1	18	1	2	1	1	0	1	
19	17	1	1	1	1	0	1	17	1	1	1	1	0	1	
20	28	1	2	1	1	0	1	28	1	2	1	1	0	1	
21	33	1	2	1	1	0	1	33	1	2	1	1	0	1	
22	22	1	1	1	1	0	0	22	1	1	1	1	0	0	
23	10	1	2	2	1	1	0	10	1	2	2	1	1	0	
24	14	1	3	2	1	1	0	14	1	3	2	1	1	0	
25	15	1	2	2	1	1	0	15	1	2	2	1	1	0	
26	17	1	2	1	1	0	0	17	1	2	1	1	0	0	
27	18	1	2	1	1	0	0	18	1	2	2	1	1	0	comma missing/long NP as subject
28	16	1	1	1	1	0	2	16	1	2	2	1	0	2	“L1” tagged as VBZ
29	17	1	3	1	1	0	0	17	1	3	1	1	0	0	
30	19	1	2	2	1	1	0	19	1	2	2	1	1	0	
31	13	1	1	1	1	0	1	13	1	1	1	1	0	1	
32	25	1	3	3	1	2	0	25	1	3	3	1	2	0	
33	27	1	3	3	2	1	1	27	1	3	3	1	0	1	cannot parse T-Unit (“therefore” inhibits)
34	35	1	2	1	1	0	1	35	1	2	1	1	0	1	
35	16	1	2	2	1	1	0	16	1	2	2	1	1	0	
36	43	1	3	3	1	2	1	43	1	3	3	1	2	1	
37	22	1	3	2	1	1	0	22	1	3	2	1	1	0	
38	26	1	4	3	1	2	0	26	1	4	3	1	2	0	
39	20	1	2	1	1	0	0	20	1	2	0	1	0	0	main clause not parsed, long subject without comma, VBP tagged as VB
40	33	1	3	2	1	1	1	33	1	3	2	1	1	1	
41	35	1	3	3	2	1	1	35	1	4	3	2	1	1	stranded “to” → coordinated VP parsed as 2 VP
42	19	1	2	2	1	1	0	19	1	2	2	1	1	0	
43	27	1	5	3	1	2	1	27	1	5	3	1	2	1	
44	31	1	6	3	1	2	1	31	1	6	3	1	2	1	
45	21	1	3	2	1	1	0	21	1	3	2	1	1	0	
46	29	1	3	2	1	1	1	29	1	3	2	1	1	1	
47	24	1	1	1	1	0	1	24	1	1	1	1	0	1	
48	50	1	3	3	1	2	2	50	1	3	3	1	2	2	
49	11	1	1	1	1	0	0	11	1	1	1	1	0	0	
50	30	1	3	2	1	1	1	30	1	3	2	1	1	1	

Comparison Sample 1 / PS text (page 2/2)															
	Manual_annotation							L2SCA_annotation							Comments
Sent.	W	S	VP	C	T	DC	CP	W	S	VP	C	T	DC	CP	
51	36	1	3	3	2	1	0	36	1	4	3	1	1	1	example list misleads parser/tagging
52	19	1	3	1	1	0	0	19	1	3	1	1	0	0	
53	21	1	1	1	1	0	3	21	1	1	1	1	0	3	
54	19	1	2	2	1	1	1	19	1	2	2	1	1	1	
55	23	1	1	1	1	0	1	23	1	1	1	1	0	1	
56	10	1	2	1	1	0	0	10	1	2	1	1	0	0	
57	33	1	4	2	1	1	2	33	1	4	2	1	1	2	
58	24	1	2	1	1	0	2	24	1	2	1	1	0	2	
59	37	1	3	2	2	0	1	37	1	3	2	2	0	1	
60	34	1	2	2	1	1	2	34	1	2	1	1	0	1	"therefore" disturbs correct parsing
61	25	1	3	3	2	1	1	25	1	3	2	2	0	1	VBP tagged as VB
62	20	1	1	1	1	0	1	20	1	1	1	1	0	1	
63	41	1	2	2	1	1	1	41	1	2	2	1	1	1	
64	15	1	1	1	1	0	0	15	1	1	1	1	0	0	
65	35	1	5	2	1	1	1	35	1	6	2	1	1	1	"read" tagged as VBD instead of VB
66	31	1	2	2	1	1	1	31	1	2	2	1	1	1	
67	35	1	6	4	2	2	1	35	1	6	4	2	2	1	
68	32	1	2	1	1	0	0	32	1	2	1	1	0	0	
69	19	1	2	1	1	0	0	19	1	2	1	1	0	0	
70	21	1	3	2	1	1	0	21	1	3	2	1	1	0	
71	14	1	2	1	1	0	0	14	1	2	1	1	0	0	
72	27	1	3	3	1	2	0	27	1	3	3	1	2	0	
73	20	1	2	1	1	0	1	20	1	2	1	1	0	1	
74	23	1	2	2	1	1	2	23	1	2	2	1	1	2	
75	39	1	8	6	3	3	1	39	1	8	6	3	3	0	"appropriate" tagged as NN not JJ
76	19	1	1	1	1	0	1	19	1	1	1	1	0	1	
77	14	1	1	1	1	0	0	14	1	1	1	1	0	0	
78	11	1	1	1	1	0	1	11	1	1	1	1	0	1	
79	26	1	1	1	1	0	1	26	1	1	1	1	0	1	
80	18	1	3	2	2	0	0	18	1	3	2	2	0	0	
81	33	1	3	3	2	1	0	33	1	3	3	2	1	0	
82	16	1	1	1	1	0	0	16	1	1	1	1	0	0	

Summary for PS paper of sample 1							
W	S	VP	C	T	DC	CP	Manual Annotations
1989	82	202	156	98	58	52	

W	S	VP	C	T	DC	CP	L2SCA Annotations
1989	82	208	153	96	53	51	

W	S	VP	C	T	DC	CP	Identical (found by both annotation methods)
1989	82	202	151	96	52	50	

Table 11: Comparison of annotations for SE paper of sample 1

Comparison Sample 1 / SE text (page 1/2)															
	Manual_annotation							**L2SCA_annotation**							**Comments**
Sent.	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	
1	9	1	1	1	1	0	0	9	1	1	1	1	0	0	
2	32	1	2	2	1	1	3	32	1	2	2	1	1	2	typo “betweendifferent” disturbs correct parsing
3	15	1	1	1	1	0	2	15	1	1	1	1	0	2	
4	18	1	3	1	1	0	1	18	1	3	1	1	0	1	
5	46	1	4	2	1	1	2	46	1	4	2	1	1	2	
6	16	1	2	2	1	1	0	16	1	2	2	1	1	0	
7	37	1	4	2	1	1	1	37	1	4	2	1	1	1	
8	34	1	3	2	2	0	1	34	1	5	2	2	0	1	“language” tagged as VP/ “however” inhibits correct parsing
9	27	1	2	2	2	0	0	27	1	2	2	2	0	0	
10	34	1	2	2	1	1	1	34	1	2	2	1	1	1	
11	24	1	1	1	1	0	1	24	1	1	1	1	0	1	
12	17	1	2	1	1	0	0	17	1	2	1	1	0	0	
13	48	1	5	2	1	1	4	48	1	5	1	1	0	4	long subject of MC disturbs parsing
14	11	1	2	1	1	0	1	11	1	2	1	1	0	1	
15	25	1	3	3	1	2	1	25	1	3	3	1	2	1	
16	20	1	2	1	1	0	0	20	1	2	1	1	0	0	
17	25	1	2	2	1	1	1	25	1	2	2	1	1	1	
18	31	1	2	1	1	1	1	31	1	2	1	1	0	1	DC not parsed
19	40	1	3	3	2	1	1	40	1	3	3	1	1	0	“hence” inhibits parsing of T-Unit
20	18	1	2	1	1	0	1	18	1	2	1	1	0	1	
21	28	1	3	2	2	0	0	28	1	2	2	2	0	0	“to process” tagged as NN
22	24	1	2	2	1	1	1	24	1	2	2	1	1	1	
23	41	1	3	2	1	1	1	41	1	4	2	1	0	1	DC not parsed, “to” tagged as VP
24	42	1	5	4	1	3	0	42	1	6	5	2	3	0	“aptidude” tagged as VP
25	12	1	1	1	1	0	1	12	1	1	1	1	0	1	
26	21	1	3	1	1	0	0	21	1	3	1	1	0	0	
27	21	1	4	2	1	1	0	21	1	4	3	1	2	0	“multilingualism” tagged as VP
28	9	1	1	1	1	0	0	9	1	1	1	1	0	0	
29	24	1	2	2	1	1	1	24	1	2	2	1	1	1	
30	9	1	1	1	1	0	1	9	1	1	0	1	0	1	“comprised” tagged as VBN → however
31	21	1	3	2	1	1	0	21	1	3	2	1	1	0	
32	12	1	2	1	1	0	2	12	1	2	1	1	0	2	
33	33	1	2	2	1	1	0	33	1	2	2	1	1	0	
34	27	1	4	2	1	1	0	27	1	4	2	1	1	0	
35	20	1	3	1	1	0	1	20	1	3	1	1	0	1	
36	16	1	3	2	1	1	0	16	1	3	2	1	1	0	
37	16	1	2	1	1	0	0	16	1	2	1	1	0	0	
38	31	1	2	2	1	1	0	31	1	2	2	1	1	0	
39	30	1	3	2	1	1	2	30	1	3	2	1	1	2	
40	19	1	1	1	1	0	1	19	1	1	1	1	0	1	
41	43	1	3	3	1	2	1	43	1	3	3	1	2	1	
42	12	1	1	1	1	0	0	12	1	1	1	1	0	0	
43	18	1	2	2	1	1	0	18	1	2	2	1	1	0	
44	27	1	3	1	1	0	1	27	1	3	1	1	0	1	
45	21	1	4	2	1	1	0	21	1	4	2	1	1	0	
46	17	1	2	2	1	1	0	17	1	2	2	1	1	0	
47	32	1	3	2	1	1	3	32	1	3	3	1	2	3	subject missing in one VP
48	37	1	5	4	1	3	0	37	1	5	4	1	3	0	
49	12	1	2	2	1	1	0	12	1	2	2	1	1	0	
50	25	1	2	2	1	1	1	25	1	3	2	1	1	1	stranded VP/unnatural → not detectable

Comparison Sample 1 / SE text (page 2/2)															
	Manual_annotation							L2SCA_annotation							Comments
Sent.	W	S	VP	C	T	DC	CP	W	S	VP	C	T	DC	CP	
51	17	1	3	2	1	1	2	17	1	3	2	1	1	2	
52	21	1	3	2	2	0	0	21	1	3	2	1	0	0	two main clauses acc. To def.
53	29	1	4	3	1	2	0	29	1	4	3	1	2	0	
54	31	1	2	2	2	0	0	31	1	2	2	2	0	0	
55	38	1	7	4	2	2	0	38	1	7	4	2	2	0	
56	25	1	2	2	1	1	1	25	1	2	2	1	1	1	
57	26	1	4	2	1	1	1	26	1	4	2	1	1	1	
58	39	1	2	2	1	1	2	39	1	3	2	1	1	2	“regarding” tagged as VP not PP
59	22	1	2	1	1	0	2	22	1	2	1	1	0	2	
60	19	1	4	3	1	2	0	19	1	4	3	1	2	0	
61	10	1	1	1	1	0	0	10	1	1	1	1	0	0	
62	9	1	1	1	1	0	0	9	1	1	1	1	0	0	
63	36	1	3	3	1	2	2	36	1	3	3	1	2	2	
64	42	1	3	3	1	2	1	42	1	3	3	1	2	1	
65	18	1	1	1	1	0	0	18	1	1	1	1	0	0	
66	16	1	2	2	1	1	0	16	1	2	2	1	1	0	
67	30	1	4	3	1	2	0	30	1	4	3	1	2	0	
68	25	1	3	3	1	2	0	25	1	3	3	1	2	0	
69	24	1	2	2	1	1	1	24	1	3	3	1	0	1	“aptidude” tagged as VP
70	7	1	1	1	1	0	0	7	1	1	1	1	0	0	
71	36	1	1	1	1	0	2	36	1	1	1	1	0	2	
72	38	1	4	2	1	1	0	38	1	4	2	1	1	0	
73	25	1	2	2	1	1	0	25	1	2	2	1	1	0	
74	28	1	2	2	1	1	1	28	1	2	2	1	0	1	“however” disturbs correct parsing
75	19	1	1	1	1	0	0	19	1	1	1	1	0	0	
76	29	1	3	2	1	1	1	29	1	3	2	1	1	1	
77	37	1	3	2	1	1	0	37	1	3	2	1	0	0	long subject of DC disturbs parsing
78	26	1	2	2	1	1	0	26	1	2	2	1	1	0	
79	43	1	4	4	2	2	1	43	1	4	4	2	2	1	

Summary for SE paper of sample 1							
W	S	VP	C	T	DC	CP	Manual Annotations
1992	79	201	149	87	63	56	

W	S	VP	C	T	DC	CP	L2SCA Annotations
1992	79	207	151	86	59	54	

W	S	VP	C	T	DC	CP	Identical (found by both annotation methods)
1992	79	200	147	85	57	54	

Table 12: Comparison of annotations for PS paper of sample 28

Comparison Sample 28 / PS text (page 1/2)															
	Manual_annotation							L2SCA_annotation							Comments
Sent.	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	
1	29	1	1	1	1	0	1	29	1	1	1	1	0	1	
2	31	1	2	2	1	1	1	31	1	2	2	1	1	1	
3	28	1	2	2	1	1	1	28	1	2	2	1	1	1	
4	19	1	2	2	1	1	1	19	1	2	2	1	1	1	
5	52	1	4	4	1	3	1	52	1	4	4	1	3	1	
6	42	1	1	1	1	0	1	42	1	1	1	1	0	1	
7	40	1	2	1	1	0	2	40	1	2	1	1	0	1	“but also” not detected as a CC
8	15	1	1	1	1	0	0	15	1	1	1	1	0	0	
9	47	1	5	3	1	2	0	47	1	5	3	1	2	0	
10	33	1	2	2	1	1	1	33	1	1	1	1	0	1	“formulated” parsed as VBN not VBD
11	30	1	2	2	1	1	2	30	1	2	2	1	1	2	
12	16	1	2	1	1	0	1	16	1	2	1	1	0	1	
13	35	1	3	1	1	0	1	35	1	3	1	1	0	1	
14	36	1	2	2	1	1	1	36	1	2	2	1	1	1	
15	27	1	4	3	1	2	1	27	1	4	3	1	2	1	
16	29	1	2	2	1	1	1	29	1	2	2	1	1	1	
17	34	1	2	2	1	1	2	34	1	2	2	1	1	2	
18	32	1	3	1	1	0	1	32	1	3	1	1	0	1	
19	39	1	3	2	1	1	1	39	1	3	2	1	1	1	
20	17	1	2	1	1	0	0	17	1	2	1	1	0	0	
21	21	1	3	2	1	1	0	21	1	3	2	1	1	0	
22	47	1	5	4	1	3	0	47	1	5	4	1	3	0	
23	27	1	2	2	2	0	2	27	1	2	2	1	0	2	“however” disturbs parsing
24	19	1	2	1	1	0	0	19	1	2	1	1	0	0	
25	25	1	2	1	1	0	1	25	1	3	1	1	0	1	“to rhoticity” tagged as VP
26	25	1	2	1	1	0	1	25	1	2	1	1	0	1	
27	35	1	5	4	1	3	2	35	1	5	4	1	3	2	
28	55	1	6	3	1	2	0	55	1	5	3	1	2	0	“leading” tagged as JJ
29	42	1	2	2	2	0	1	42	1	2	2	1	0	1	“however” disturbs parsing
30	17	1	2	1	1	0	0	17	1	3	1	1	0	0	“rhoticity” parsed as VP
31	23	1	3	3	1	2	0	23	1	3	3	1	2	0	
32	32	1	3	1	1	0	1	32	1	3	1	1	0	1	
33	13	1	1	1	1	0	0	13	1	1	1	1	0	0	
34	49	1	5	3	1	2	0	49	1	4	3	1	2	0	one VP not parsed, unclear why
35	24	1	2	1	1	0	0	24	1	1	1	1	0	0	“aggravating” tagged as JJ
36	32	1	3	2	1	1	1	32	1	3	2	1	1	1	
37	49	1	3	2	2	0	2	49	1	3	2	2	0	2	
38	48	1	5	3	1	2	2	48	1	5	3	1	2	2	
39	53	1	4	2	2	0	1	53	1	4	2	2	0	1	
40	39	1	5	2	1	1	2	39	1	5	2	1	1	2	
41	46	1	5	3	1	2	1	46	1	5	3	1	2	1	
42	24	1	3	3	2	1	1	24	1	3	3	2	1	1	
43	31	1	4	2	1	1	2	31	1	4	2	1	1	2	
44	47	1	2	2	1	1	2	47	1	2	2	1	1	2	
45	29	1	2	2	1	1	1	29	1	2	2	1	1	1	
46	49	1	4	3	1	2	1	49	1	4	4	1	3	1	long subject VP disturbs parsing
47	37	1	4	3	1	2	1	37	1	4	3	1	2	1	
48	22	1	3	2	1	1	0	22	1	3	2	1	1	0	
49	59	1	8	6	1	4	1	59	1	8	6	1	3	1	DC not parsed correctly
50	32	1	3	1	1	0	2	32	1	3	2	1	1	2	“measures” tagged as VP → leads to parsing a DC

Comparison Sample 28 / PS text (page 2/2)

	Manual_annotation							L2SCA_annotation							Comments
Sent.	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	
51	19	1	1	1	1	0	2	19	1	1	1	1	0	2	
52	28	1	3	1	1	0	0	28	1	3	1	1	0	0	
53	56	1	5	4	1	3	1	56	1	5	4	1	3	1	
54	19	1	2	1	1	0	0	19	1	2	1	1	0	0	
55	30	1	3	2	1	1	0	30	1	3	2	1	1	0	
56	18	1	2	1	1	0	0	18	1	2	1	1	0	0	
57	27	1	2	1	1	0	1	27	1	2	2	1	1	1	“whichever” parses DC

Summary of annotations for PS paper of sample 28

W	S	VP	C	T	DC	CP	
1879	57	168	115	62	52	52	**Manual Annotations**

W	S	VP	C	T	DC	CP	
1879	57	166	117	60	53	51	**L2SCA Annotations**

W	S	VP	C	T	DC	CP	
1879	57	164	114	60	50	51	**Identical (found by both annotation methods)**

Table 13: Comparison of annotations for SE paper of sample 28

Comparison Sample 28 / SE text (page 1/2)															
	Manual_annotation							**L2SCA_annotation**							**Comments**
Sent.	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	
1	22	1	1	1	1	0	1	22	1	1	1	1	0	1	
2	28	1	3	2	1	1	1	28	1	3	2	1	1	1	
3	42	1	1	1	1	0	3	42	1	1	1	1	0	3	
4	19	1	3	1	1	0	0	19	1	3	1	1	0	0	
5	40	1	4	2	1	1	1	40	1	4	2	1	1	1	
6	25	1	1	1	1	0	2	25	1	2	1	1	0	2	“rather” disturbs tagging of whole VP
7	39	1	5	3	1	2	0	39	1	5	3	1	2	0	
8	23	1	3	2	1	1	2	23	1	2	1	1	0	2	
9	29	1	3	3	1	2	1	29	1	3	3	1	2	1	
10	26	1	2	1	1	0	1	26	1	2	1	1	0	1	
11	24	1	3	3	1	2	0	24	1	3	3	1	2	0	
12	14	1	3	3	1	2	0	14	1	3	3	1	2	0	
13	24	1	2	2	1	1	1	24	1	2	2	1	1	1	
14	28	1	2	2	1	1	1	28	1	2	1	1	0	1	dashes disturb parsing of DC
15	25	1	3	2	1	1	0	25	1	3	3	1	1	0	“as well as” has coordinating function,
16	22	1	2	1	1	0	2	22	1	2	1	1	0	2	subject is missing for 2nd DC
17	27	1	3	3	1	2	1	27	1	3	3	1	2	1	
18	20	1	1	1	1	0	2	20	1	1	1	1	0	2	
19	28	1	3	3	1	2	0	28	1	3	3	1	2	0	
20	33	1	3	2	2	0	2	33	1	3	2	2	0	2	
21	39	1	4	4	1	3	1	39	1	4	4	1	3	1	
22	12	1	1	1	1	0	0	12	1	1	1	1	0	0	
23	24	1	3	1	1	0	2	24	1	3	1	1	0	2	
24	28	1	4	2	1	1	0	28	1	3	1	1	0	0	VP “functions” tagged as NP
25	19	1	1	1	1	0	0	19	1	1	1	1	0	0	
26	18	1	2	1	1	0	1	18	1	2	1	1	0	1	
27	30	1	4	1	1	0	2	30	1	5	2	1	1	2	list of VPs are misleading
28	30	1	4	1	1	0	1	30	1	4	1	1	0	0	“but” not detected, VP count is correct
29	28	1	2	1	1	0	2	28	1	2	1	1	0	2	
30	30	1	2	1	1	0	2	30	1	2	1	1	0	2	
31	16	1	1	1	1	0	1	16	1	1	1	1	0	1	
32	24	1	4	4	1	3	0	24	1	4	4	1	3	0	
33	29	1	4	3	1	2	1	29	1	4	3	1	2	1	
34	24	1	2	2	1	1	2	24	1	2	2	1	1	2	
35	33	1	4	4	1	3	2	33	1	4	4	1	3	2	
36	20	1	1	1	1	0	1	20	1	1	1	1	0	1	
37	46	1	4	3	2	1	3	46	1	4	3	1	1	3	2nd T-unit is in dashes → not detected
38	21	1	2	2	1	1	0	21	1	2	2	1	1	0	
39	9	1	1	1	1	0	0	9	1	1	1	1	0	0	
40	19	1	1	1	1	0	0	19	1	1	1	1	0	0	
41	26	1	2	1	1	0	0	26	1	2	1	1	0	0	
42	31	1	2	2	1	1	3	31	1	2	2	1	1	3	
43	52	1	2	1	1	0	2	52	1	2	1	1	0	2	
44	33	1	2	2	1	1	1	33	1	2	2	1	1	1	
45	38	1	4	3	1	2	0	38	1	4	3	1	2	0	
46	18	1	2	1	1	0	1	18	1	2	1	1	0	1	
47	26	1	2	2	1	1	0	26	1	2	2	1	1	0	
48	27	1	2	2	1	1	1	27	1	2	2	1	1	1	
49	31	1	2	2	1	1	1	31	1	2	2	1	1	1	
50	40	1	6	6	6	0	0	40	1	6	5	4	1	0	examples in the sentence disturbs parsing

Comparison Sample 28 / SE text (page 2/2)

	Manual_annotation							L2SCA_annotation							Comments
Sent.	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	
51	43	1	3	3	3	0	1	43	1	3	3	1	0	1	examples in the sentence disturbs correct parsing
52	30	1	3	3	1	2	1	30	1	3	3	1	2	1	
53	37	1	3	3	2	1	0	37	1	3	3	1	0	0	DC not detected
54	25	1	4	1	1	0	0	25	1	4	1	1	0	0	
55	21	1	2	1	1	0	0	21	1	2	1	1	0	0	
56	14	1	2	2	1	1	0	14	1	2	2	1	1	0	
57	28	1	4	3	1	2	0	28	1	4	3	1	2	0	
58	28	1	2	2	1	1	0	28	1	2	2	1	1	0	
59	20	1	3	2	1	1	0	20	1	3	2	1	1	0	
60	21	1	1	1	1	0	1	21	1	1	1	1	0	1	
61	19	1	2	2	1	1	0	19	1	3	2	1	1	0	“charisma” tagged as VP
62	27	1	2	2	1	1	1	27	1	3	2	1	1	1	“fame” tagged as VP
63	14	1	1	1	1	0	1	14	1	1	1	1	0	1	
64	22	1	2	2	1	1	2	22	1	2	2	1	1	2	
65	35	1	3	3	1	2	1	35	1	3	3	1	2	1	
66	38	1	5	4	2	2	0	38	1	5	4	2	2	0	
67	27	1	2	1	1	0	0	27	1	2	1	1	0	0	
68	26	1	1	1	1	0	0	26	1	1	1	1	0	0	
69	19	1	2	2	1	1	1	19	1	2	2	1	1	1	
70	20	1	2	1	1	0	2	20	1	2	1	1	0	2	
71	37	1	2	2	1	1	1	37	1	3	2	1	1	1	“Nasser” tagged as VP
72	21	1	2	1	1	0	1	21	1	2	1	1	0	1	
73	27	1	2	1	1	0	1	27	1	2	1	1	0	1	
74	33	1	5	3	1	2	2	33	1	5	3	1	2	2	

Summary of annotations for SE paper of sample 28

W	**S**	**VP**	**C**	**T**	**DC**	**CP**	
1991	74	188	144	85	59	67	**Manual Annotations**

W	**S**	**VP**	**C**	**T**	**DC**	**CP**	
1991	74	191	142	79	57	66	**L2SCA Annotations**

W	**S**	**VP**	**C**	**T**	**DC**	**CP**	
1991	74	186	140	79	55	66	**Identical (found by both annotation methods)**

Table 14: Comparison of annotations for PS paper of sample 57

Comparison Sample 57 / PS text (page 1/2)															
	Manual_annotation							**L2SCA_annotation**							**Comments**
Sent.	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	
1	11	1	2	2	2	0	0	11	1	1	1	1	0	1	question not parsed as MC
2	22	1	2	1	1	1	1	22	1	2	1	1	0	1	question not parsed as DC
3	31	1	2	1	1	0	2	31	1	2	1	1	0	2	
4	20	1	3	2	1	1	1	20	1	3	1	1	0	1	DC not parsed
5	26	1	2	1	1	0	1	26	1	2	1	1	0	1	
6	9	1	1	0	0	0	0	9	1	2	0	1	0	0	no grammatically correct sentence,
7	21	1	2	1	1	0	0	21	1	2	1	1	0	0	but sentence according to definition
8	12	1	2	1	1	0	0	12	1	2	1	1	0	0	
9	43	1	6	4	1	3	2	43	1	6	4	1	3	2	
10	10	1	1	1	1	0	0	10	1	1	1	1	0	0	
11	18	1	5	2	1	1	0	18	1	5	2	1	1	0	
12	35	1	8	6	2	4	0	35	1	7	6	2	4	0	“works” tagged as NNS
13	23	1	4	3	1	2	0	23	1	4	3	1	2	0	
14	30	1	4	3	1	2	0	30	1	4	3	1	2	0	
15	32	1	4	4	1	3	2	32	1	4	4	1	3	2	
16	25	1	5	4	1	3	2	25	1	5	4	1	3	2	
17	18	1	3	2	1	1	0	18	1	4	3	1	2	0	“it” tagged wrong → disturbs DC-parsing
18	20	1	4	3	2	1	0	20	1	4	2	1	2	0	questions in text → disturb parsing
19	29	1	1	1	1	0	1	29	1	1	1	1	0	1	
20	23	1	2	1	1	0	1	23	1	2	1	1	0	1	
21	19	1	2	1	1	0	1	19	1	2	1	1	0	1	
22	36	1	3	3	2	1	0	36	1	3	3	2	1	0	
23	29	1	3	3	1	2	2	29	1	4	2	1	1	2	“have” & “produce” tagged VB, not VBP
24	19	1	2	1	1	0	2	19	1	2	1	1	0	2	
25	29	1	3	3	1	2	1	29	1	3	3	1	2	1	
26	8	1	2	1	1	0	0	8	1	2	1	1	0	0	
27	33	1	5	3	1	2	1	33	1	5	3	1	2	1	
28	22	1	3	2	1	1	0	22	1	3	2	1	1	0	
29	32	1	3	3	1	2	1	32	1	3	3	1	2	1	
30	16	1	4	2	1	1	0	16	1	4	2	1	1	0	
31	13	1	3	2	1	1	0	13	1	3	2	1	1	0	
32	21	1	4	3	1	2	1	21	1	4	3	1	2	1	
33	58	1	7	6	1	5	1	58	1	7	6	1	5	1	
34	21	1	3	2	1	1	0	21	1	3	2	1	1	0	
35	15	1	3	2	1	1	0	15	1	3	2	1	1	0	
36	42	1	6	5	1	4	1	42	1	6	5	1	4	1	
37	31	1	4	3	2	1	0	31	1	5	4	1	1	0	pl. noun “lives” tagged as VP
38	54	1	9	4	1	3	0	54	1	10	6	1	5	0	“hypothesis” tagged as VP
39	54	1	7	6	2	4	2	54	1	7	6	1	4	1	“merely” inhibits CP parsing,
40	15	1	1	1	1	0	0	15	1	1	1	1	0	0	MC not dedected
41	40	1	5	4	2	2	0	40	1	5	4	2	1	0	2nd DC not detected
42	28	1	5	3	2	1	1	28	1	5	4	2	2	0	“at that time” disturbs parsing
43	12	1	2	1	1	0	0	12	1	2	1	1	0	0	
44	17	1	2	2	1	1	1	17	1	2	2	1	1	1	
45	59	1	11	6	2	4	1	59	1	13	7	2	5	1	“phonology” tagged as VP
46	45	1	6	5	2	3	1	45	1	7	5	2	3	1	“also” disturbs coordinated VP detection
47	23	1	3	3	1	2	0	23	1	3	3	1	2	0	
48	27	1	6	3	1	2	1	27	1	4	3	1	2	1	unnatural word order disturbs tagging
49	18	1	3	3	1	2	0	18	1	3	3	1	2	0	
50	12	1	2	2	1	1	0	12	1	2	2	1	1	0	

Comparison Sample 57 / PS text (page 2/2)															
	Manual_annotation							**L2SCA_annotation**							**Comments**
Sent.	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	
51	20	1	2	1	1	0	1	20	1	2	1	1	0	1	
52	28	1	4	3	2	1	0	28	1	4	3	2	1	0	
53	55	1	6	6	2	4	2	55	1	6	6	2	4	2	
54	48	1	7	6	2	4	0	48	1	7	6	2	4	0	
55	28	1	5	2	1	1	1	28	1	5	2	1	1	1	
56	36	1	6	5	3	2	1	36	1	6	5	3	2	1	
57	19	1	3	2	2	0	0	19	1	3	2	2	0	0	
58	46	1	6	5	1	4	2	46	1	6	5	1	4	2	
59	28	1	4	4	1	3	1	28	1	4	4	1	3	1	
60	48	1	6	6	2	4	0	48	1	6	6	1	5	0	wrong use of “that” parses DC
61	49	1	7	5	2	3	0	49	1	7	5	1	2	0	unnatural clause order disturbs parsing
62	38	1	7	5	2	2	1	38	1	7	5	1	2	1	coordination between MCs not detected
63	25	1	3	2	1	1	2	25	1	3	2	1	1	2	
64	31	1	5	4	1	3	0	31	1	5	4	1	3	0	
65	39	1	6	6	1	5	1	39	1	6	6	1	5	1	
66	12	1	1	1	1	0	0	12	1	1	1	1	0	0	
67	24	1	3	3	1	2	1	24	1	3	3	1	2	1	
68	29	1	4	4	1	3	0	29	1	4	4	1	3	0	
69	21	1	2	1	1	0	0	21	1	2	1	1	0	0	
70	23	1	3	2	2	0	0	23	1	3	2	2	0	0	
71	50	1	6	5	2	3	1	50	1	6	5	2	3	1	

Summary of annotations for PS paper of sample 57							
W	**S**	**VP**	**C**	**T**	**DC**	**CP**	**Manual Annotations**
2003	71	281	209	91	118	45	

W	**S**	**VP**	**C**	**T**	**DC**	**CP**	**L2SCA Annotations**
2003	71	285	211	85	120	44	

W	**S**	**VP**	**C**	**T**	**DC**	**CP**	**Identical (found by both annotation methods)**
2003	71	277	205	84	113	43	

Table 15: Comparison of annotations for SE paper of sample 57

Comparison Sample 57 / SE text (page 1/2)															
	Manual_annotation							**L2SCA_annotation**							**Comments**
Sent.	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	**W**	**S**	**VP**	**C**	**T**	**DC**	**CP**	
1	12	1	2	2	1	1	0	12	1	2	2	1	1	0	
2	16	1	1	1	1	0	1	16	1	1	1	1	0	1	
3	9	1	2	2	1	1	0	9	1	2	2	1	1	0	
4	15	1	1	1	1	0	1	15	1	1	1	1	0	1	
5	62	1	7	5	4	1	2	62	1	7	5	1	4	2	thesis statements disturb parsing
6	25	1	6	5	1	4	0	25	1	6	5	1	4	0	
7	29	1	3	3	2	1	0	29	1	2	2	2	0	0	VP “summarizes” tagged as NN
8	25	1	1	1	1	0	2	25	1	1	1	1	0	2	
9	33	1	2	2	1	1	1	33	1	2	1	1	0	1	VP “amount” tagged as NN
10	32	1	3	3	1	2	1	32	1	3	3	1	2	1	
11	22	1	3	3	1	2	1	22	1	4	3	1	2	1	DC is incorrectly parsed → more VPs
12	15	1	2	2	1	1	2	15	1	2	1	1	0	2	not clear why DC is not parsed
13	43	1	4	2	2	0	2	43	1	4	2	2	0	2	
14	36	1	6	4	2	2	0	36	1	7	5	2	3	0	”features” tagged as VBZ
15	34	1	4	2	1	1	2	34	1	4	1	1	1	2	VP “to learn a...” in list disturbs parsing
16	28	1	3	1	1	0	1	28	1	3	1	1	0	1	
17	18	1	2	2	1	1	0	18	1	2	2	1	1	0	
18	11	1	1	1	1	0	0	11	1	1	1	1	0	0	
19	29	1	3	2	1	1	1	29	1	3	2	1	1	1	
20	16	1	2	2	1	1	0	16	1	2	2	1	1	0	
21	27	1	3	3	1	2	0	27	1	3	3	1	2	0	
22	14	1	1	1	1	0	0	14	1	1	1	1	0	0	
23	25	1	1	1	1	0	1	25	1	1	1	1	0	1	
24	19	1	2	2	1	1	0	19	1	2	2	1	1	0	
25	16	1	2	1	1	0	1	16	1	2	1	1	0	1	
26	21	1	1	1	1	0	2	21	1	1	1	1	0	2	
27	38	1	8	4	1	3	1	38	1	8	4	1	3	1	
28	41	1	4	4	1	3	2	41	1	4	4	1	3	2	
29	16	1	4	2	1	1	0	16	1	4	2	1	1	0	
30	28	1	3	3	1	2	0	28	1	3	3	1	2	0	
31	14	1	1	1	1	0	0	14	1	1	1	1	0	0	
32	10	1	2	1	1	0	0	10	1	2	1	1	0	0	
33	14	1	1	1	1	0	0	14	1	1	1	1	0	0	
34	18	1	3	2	1	1	0	18	1	3	2	1	1	0	
35	28	1	4	3	1	2	0	28	1	4	3	1	2	0	
36	12	1	3	2	1	1	0	12	1	3	2	1	1	0	
37	21	1	3	2	2	0	0	21	1	3	2	1	1	0	“however” inhibits MC detection
38	17	1	1	1	1	0	1	17	1	1	1	1	0	1	
39	26	1	3	3	1	2	1	26	1	3	3	1	2	1	
40	25	1	4	2	1	1	1	25	1	4	2	1	1	1	
41	15	1	1	1	1	0	1	15	1	1	1	1	0	1	
42	16	1	1	1	1	0	1	16	1	1	1	1	0	1	
43	18	1	3	2	1	1	0	18	1	3	2	1	1	0	
44	8	1	1	1	1	0	0	8	1	1	1	1	0	0	
45	19	1	2	1	1	0	2	19	1	2	1	1	0	2	
46	18	1	2	1	1	0	1	18	1	2	1	1	0	1	
47	19	1	2	1	1	0	0	19	1	2	1	1	0	0	
48	32	1	3	2	1	1	1	32	1	3	2	1	1	1	
49	32	1	3	3	1	2	2	32	1	3	3	1	2	2	
50	21	1	3	1	1	0	3	21	1	3	1	1	0	3	

Comparison Sample 57 / SE text (page 2/2)															
	Manual_annotation							L2SCA_annotation							Comments
Sent.	W	S	VP	C	T	DC	CP	W	S	VP	C	T	DC	CP	
51	19	1	2	2	1	1	1	19	1	2	2	1	1	1	
52	47	1	8	5	5	0	0	47	1	9	5	5	0	0	"meaning" not tagged as NN
53	14	1	1	1	1	0	1	14	1	1	1	1	0	1	
54	13	1	2	1	1	0	1	13	1	2	1	1	0	1	
55	15	1	1	1	1	0	0	15	1	1	1	1	0	0	
56	28	1	2	2	1	1	1	28	1	2	2	1	1	1	
57	16	1	2	2	1	1	0	16	1	2	2	1	1	0	
58	25	1	2	1	1	0	0	25	1	2	2	1	1	0	"empirically" tagged as NN, "even with" parses DC, unnatural long subject
59	23	1	1	1	1	0	2	23	1	1	1	1	0	2	
60	49	1	3	1	1	0	4	49	1	3	1	1	0	4	
61	37	1	4	3	1	2	2	37	1	4	3	1	2	2	
62	16	1	3	2	1	1	1	16	1	3	2	1	1	1	
63	18	1	3	2	2	0	0	18	1	3	2	2	0	0	
64	27	1	6	5	1	4	1	27	1	6	5	1	4	1	
65	35	1	6	4	1	3	2	35	1	7	4	1	3	2	"learning styles" tagged as VP
66	16	1	2	2	1	1	0	16	1	1	1	1	0	0	"summarizes" tagged as NN
67	9	1	1	1	1	0	0	9	1	1	1	1	0	0	
68	22	1	1	1	1	0	3	22	1	1	1	1	0	3	
69	18	1	2	2	1	1	1	18	1	2	2	1	1	0	CP not detected
70	17	1	1	1	1	0	2	17	1	1	1	1	0	1	CP not detected
71	17	1	2	2	1	1	1	17	1	2	2	1	1	1	
72	28	1	2	2	1	1	2	28	1	2	2	1	1	2	
73	40	1	2	1	1	0	5	40	1	3	0	0	0	4	"results in" tagged as NNS, not VP
74	7	1	1	1	1	0	0	7	1	1	1	1	0	0	
75	20	1	2	1	1	0	1	20	1	2	1	1	0	1	
76	49	1	5	4	2	2	2	49	1	6	4	2	2	1	"perceiving" not tagged as NN
77	44	1	5	5	4	1	3	44	1	5	5	4	1	3	
78	31	1	1	1	1	0	6	31	1	2	1	1	0	6	"sensing learner" tagged as VP
79	23	1	3	2	1	1	3	23	1	3	2	1	1	2	CP not detected
80	15	1	2	2	1	1	1	15	1	2	2	1	1	1	
81	29	1	4	3	1	2	1	29	1	4	3	1	2	1	
82	22	1	5	2	2	0	1	22	1	5	2	2	0	1	
83	15	1	2	2	1	1	0	15	1	2	2	1	1	0	
84	23	1	2	2	1	1	2	23	1	2	2	1	1	2	
85	8	1	2	1	1	0	0	8	1	2	1	1	0	0	
86	22	1	2	1	1	0	1	22	1	2	1	1	0	1	

Summary of annotations for SE paper of sample 57							
W	S	VP	C	T	DC	CP	Manual Annotations
1995	86	228	172	103	69	89	

W	S	VP	C	T	DC	CP	L2SCA Annotations
1995	86	233	168	98	71	84	

W	S	VP	C	T	DC	CP	Identical (found by both annotation methods)
1995	86	226	166	98	65	84	

8 Summary of manual and L2SCA comparison

Table 16: Summary of annotations for all 6 samples

L2SCA annotations vs. manual annotation																									
Sample	**S**	**S_M**	**Delta**	**ident.**	**TU**	**T_M**	**Delta**	**ident.**	**C**	**C_M**	**Delta**	**ident.**	**DC**	**DC_M**	**Delta**	**ident.**	**VP**	**VP_M**	**Delta**	**ident.**	**CP**	**CP_M**	**Delta**	**ident.**	**W**	**W_M**	**Delta**	**ident.**
1_PS	82	82	0	82	96	98	2	96	153	156	3	151	53	58	5	52	208	202	6	202	51	52	1	50	1989	1989	0	1989
1_SE	79	79	0	79	86	87	2	85	151	149	3	147	59	63	5	57	207	201	6	200	54	56	2	54	1992	1992	0	1992
28_PS	57	57	0	57	60	62	1	60	117	115	2	114	53	52	4	50	166	168	2	164	51	52	1	51	1879	1879	0	1879
28_SE	74	74	0	74	79	85	2	79	142	144	2	140	57	59	1	55	191	188	3	186	66	67	1	66	1991	1991	0	1991
57_PS	71	71	0	71	85	91	6	84	211	209	2	205	120	118	2	113	285	281	4	277	44	45	1	43	2003	2003	0	2003
57_SE	86	86	0	86	98	103	6	98	168	172	2	166	71	69	2	65	233	228	5	226	84	89	5	84	1995	1995	0	1995
Sum	449	449	0	449	504	526	19	502	942	945	14	923	413	419	19	392	1290	1268	26	1255	350	361	11	348	11849	11849	0	11849
Mean	74.8	74.8			84.0	87.7			157.0	157.5			68.8	69.8			215.0	211.3			58.3	60.2			1974.8	1974.8		
STDev	10.3	10.3			13.8	14.3			31.3	31.4			25.9	24.3			40.8	39.4			14.5	15.9			47.2	47.2		
STDerr	4.19	4.19			5.62	5.83			12.80	12.81			10.58	9.90			16.65	16.07			5.91	6.48			19.27	19.27		
Median	76.5	76.5			85.5	89.0			152.0	152.5			58.0	61.0			207.5	201.5			52.5	54.0			1991.5	1991.5		
Min	57	57			60	62			117	115			53	52			166	168			44	45			1879	1879		
Max	86	86			98	103			211	209			120	118			285	281			84	89			2003	2003		

Table 17: Correlation data between manual and L2SCA measures

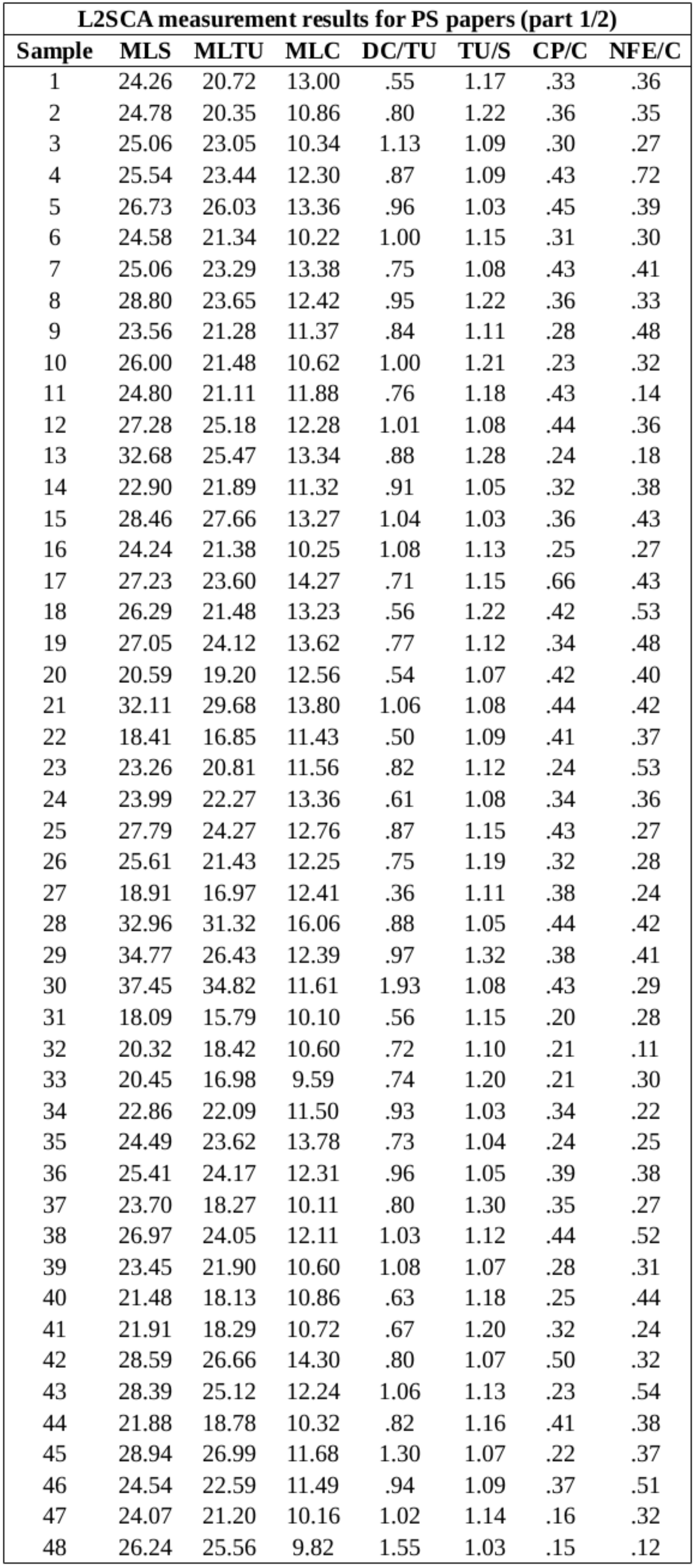
Correlation results between measures calculated from L2SCA and manual annotations														
Paper	**MLS**	**MLS_M**	**MLTU**	**MLTU_M**	**MLC**	**MLC_M**	**DC/TU**	**DC/TU_M**	**TU/S**	**TU/S_M**	**CP/C**	**CP/C_M**	**NFE/C**	**NFE/C_M**
1_PS	24.26	24.26	20.72	20.30	13.00	12.75	.552	.592	1.171	1.195	.333	.333	.359	.295
1_SE	25.22	25.22	23.16	22.90	13.19	13.37	.686	.724	1.089	1.101	.358	.376	.371	.349
28_PS	32.96	32.96	31.32	30.31	16.06	16.34	.883	.839	1.053	1.088	.436	.452	.419	.461
28_SE	26.91	26.91	25.20	23.42	14.02	13.83	.722	.694	1.068	1.149	.465	.465	.345	.306
57_PS	28.21	28.21	23.56	22.01	9.49	9.58	1.412	1.297	1.197	1.282	.209	.215	.351	.344
57_SE	23.20	23.20	20.36	19.37	11.88	11.60	.724	.670	1.140	1.198	.500	.517	.387	.326
Mean	26.79	26.79	24.05	23.05	12.94	12.91	.830	.803	1.119	1.169	.383	.393	.372	.347
STDev	3.52	3.52	4.00	3.87	2.19	2.26	.304	.255	.059	.072	.107	.109	.027	.060
STDerr	1.44	1.44	1.63	1.58	.89	.92	.124	.104	.024	.029	.043	.045	.011	.024
Median	26.06	26.06	23.36	22.45	13.10	13.06	.723	.709	1.114	1.172	.397	.414	.365	.335
Min	23.20	23.20	20.36	19.37	9.49	9.58	.552	.592	1.053	1.088	.209	.215	.345	.295
Max	32.96	32.96	31.32	30.31	16.06	16.34	1.412	1.297	1.197	1.282	.500	.517	.419	.461
Variance	12.394	12.394	15.978	15.007	4.797	5.124	.092	.065	.003	.005	.011	.012	.001	.004
Covariance	12.436		15.220		4.905		.077		.004		.012		.001	
Pearson's *r*	1		.988		.995		.993		.912		.997		.839	
p	2.2E-016***		0.0001992***		0.00003573***		0.00007556***		0.01117*		0.00001172***		0.0366*	
t (df =4)	1.34E+08		13.047		20.161		16.687		4.4589		26.686		3.0894	

Significance level: ***$p \le .001$, **$p \le .01$, *$p \le .05$

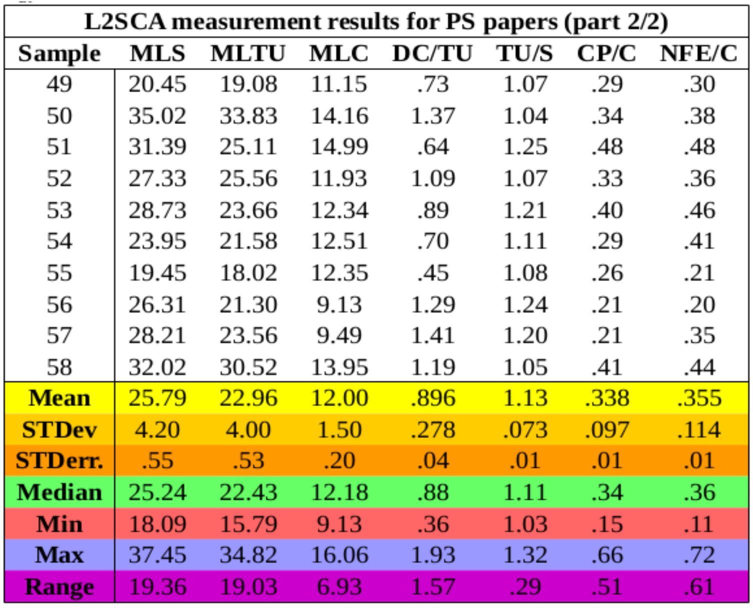
Landscape version of table 17 and 18 at the end of the appendix.

9 Measurement results

Table 18: Results for PS samples



L2SCA measurement results for PS papers (part 1/2)							
Sample	**MLS**	**MLTU**	**MLC**	**DC/TU**	**TU/S**	**CP/C**	**NFE/C**
1	24.26	20.72	13.00	.55	1.17	.33	.36
2	24.78	20.35	10.86	.80	1.22	.36	.35
3	25.06	23.05	10.34	1.13	1.09	.30	.27
4	25.54	23.44	12.30	.87	1.09	.43	.72
5	26.73	26.03	13.36	.96	1.03	.45	.39
6	24.58	21.34	10.22	1.00	1.15	.31	.30
7	25.06	23.29	13.38	.75	1.08	.43	.41
8	28.80	23.65	12.42	.95	1.22	.36	.33
9	23.56	21.28	11.37	.84	1.11	.28	.48
10	26.00	21.48	10.62	1.00	1.21	.23	.32
11	24.80	21.11	11.88	.76	1.18	.43	.14
12	27.28	25.18	12.28	1.01	1.08	.44	.36
13	32.68	25.47	13.34	.88	1.28	.24	.18
14	22.90	21.89	11.32	.91	1.05	.32	.38
15	28.46	27.66	13.27	1.04	1.03	.36	.43
16	24.24	21.38	10.25	1.08	1.13	.25	.27
17	27.23	23.60	14.27	.71	1.15	.66	.43
18	26.29	21.48	13.23	.56	1.22	.42	.53
19	27.05	24.12	13.62	.77	1.12	.34	.48
20	20.59	19.20	12.56	.54	1.07	.42	.40
21	32.11	29.68	13.80	1.06	1.08	.44	.42
22	18.41	16.85	11.43	.50	1.09	.41	.37
23	23.26	20.81	11.56	.82	1.12	.24	.53
24	23.99	22.27	13.36	.61	1.08	.34	.36
25	27.79	24.27	12.76	.87	1.15	.43	.27
26	25.61	21.43	12.25	.75	1.19	.32	.28
27	18.91	16.97	12.41	.36	1.11	.38	.24
28	32.96	31.32	16.06	.88	1.05	.44	.42
29	34.77	26.43	12.39	.97	1.32	.38	.41
30	37.45	34.82	11.61	1.93	1.08	.43	.29
31	18.09	15.79	10.10	.56	1.15	.20	.28
32	20.32	18.42	10.60	.72	1.10	.21	.11
33	20.45	16.98	9.59	.74	1.20	.21	.30
34	22.86	22.09	11.50	.93	1.03	.34	.22
35	24.49	23.62	13.78	.73	1.04	.24	.25
36	25.41	24.17	12.31	.96	1.05	.39	.38
37	23.70	18.27	10.11	.80	1.30	.35	.27
38	26.97	24.05	12.11	1.03	1.12	.44	.52
39	23.45	21.90	10.60	1.08	1.07	.28	.31
40	21.48	18.13	10.86	.63	1.18	.25	.44
41	21.91	18.29	10.72	.67	1.20	.32	.24
42	28.59	26.66	14.30	.80	1.07	.50	.32
43	28.39	25.12	12.24	1.06	1.13	.23	.54
44	21.88	18.78	10.32	.82	1.16	.41	.38
45	28.94	26.99	11.68	1.30	1.07	.22	.37
46	24.54	22.59	11.49	.94	1.09	.37	.51
47	24.07	21.20	10.16	1.02	1.14	.16	.32
48	26.24	25.56	9.82	1.55	1.03	.15	.12



L2SCA measurement results for PS papers (part 2/2)

Sample	MLS	MLTU	MLC	DC/TU	TU/S	CP/C	NFE/C
49	20.45	19.08	11.15	.73	1.07	.29	.30
50	35.02	33.83	14.16	1.37	1.04	.34	.38
51	31.39	25.11	14.99	.64	1.25	.48	.48
52	27.33	25.56	11.93	1.09	1.07	.33	.36
53	28.73	23.66	12.34	.89	1.21	.40	.46
54	23.95	21.58	12.51	.70	1.11	.29	.41
55	19.45	18.02	12.35	.45	1.08	.26	.21
56	26.31	21.30	9.13	1.29	1.24	.21	.20
57	28.21	23.56	9.49	1.41	1.20	.21	.35
58	32.02	30.52	13.95	1.19	1.05	.41	.44
Mean	25.79	22.96	12.00	.896	1.13	.338	.355
STDev	4.20	4.00	1.50	.278	.073	.097	.114
STDerr.	.55	.53	.20	.04	.01	.01	.01
Median	25.24	22.43	12.18	.88	1.11	.34	.36
Min	18.09	15.79	9.13	.36	1.03	.15	.11
Max	37.45	34.82	16.06	1.93	1.32	.66	.72
Range	19.36	19.03	6.93	1.57	.29	.51	.61

Table 19: Results for SE samples

L2SCA measurement results for SE papers (part 1/2)							
Sample	**MLS**	**MLTU**	**MLC**	**DC/TU**	**TU/S**	**CP/C**	**NFE/C**
1	25.22	23.16	13.19	.69	1.09	.36	.37
2	23.36	19.47	10.51	.82	1.20	.33	.25
3	32.35	28.66	14.23	1.03	1.13	.56	.51
4	26.63	24.65	15.13	.65	1.08	.38	.33
5	32.41	32.41	18.14	.79	1.00	.72	.59
6	34.12	29.54	12.61	1.25	1.16	.29	.46
7	25.42	23.33	14.58	.60	1.09	.52	.37
8	22.63	19.91	12.07	.60	1.14	.33	.32
9	27.05	23.28	11.99	.91	1.16	.32	.40
10	30.74	27.75	12.73	1.15	1.11	.36	.34
11	23.07	21.57	11.21	.85	1.07	.40	.34
12	26.63	22.95	13.05	.77	1.16	.35	.35
13	37.68	32.21	15.98	.87	1.17	.73	.28
14	30.03	27.15	11.52	1.38	1.11	.51	.33
15	36.13	32.57	16.42	.95	1.11	.79	.36
16	25.83	24.86	11.37	1.20	1.04	.34	.11
17	27.45	23.30	13.36	.69	1.18	.66	.27
18	30.09	24.52	14.71	.68	1.23	.46	.42
19	27.48	25.39	12.78	.96	1.08	.34	.43
20	24.02	22.16	14.99	.48	1.08	.46	.37
21	30.37	28.61	14.30	.99	1.06	.51	.36
22	18.64	18.14	11.15	.65	1.03	.40	.29
23	21.82	20.69	12.11	.69	1.05	.38	.41
24	25.95	22.93	11.95	.91	1.13	.33	.25
25	29.41	24.10	13.07	.76	1.22	.50	.33
26	24.74	21.09	11.79	.72	1.17	.33	.37
27	22.34	20.11	11.49	.76	1.11	.39	.38
28	26.91	25.20	14.02	.72	1.07	.46	.35
29	27.05	22.19	13.62	.57	1.22	.33	.30
30	26.22	25.23	10.60	1.37	1.04	.19	.34
31	20.29	19.32	11.04	.76	1.05	.34	.25
32	20.12	18.26	11.47	.59	1.10	.25	.26
33	20.25	19.29	11.19	.72	1.05	.30	.36
34	21.77	20.42	11.07	.82	1.07	.32	.26
35	21.88	21.41	15.43	.39	1.02	.35	.39
36	24.69	20.62	10.20	.96	1.20	.24	.28
37	27.64	23.98	10.87	1.12	1.15	.33	.33
38	29.94	26.39	13.83	.92	1.13	.43	.43
39	31.85	31.35	13.72	1.30	1.02	.58	.51
40	23.00	20.84	10.64	.90	1.10	.24	.38
41	24.36	22.22	10.01	1.20	1.10	.25	.38
42	31.19	28.44	14.12	1.04	1.10	.47	.39
43	26.04	22.24	13.94	.58	1.17	.46	.53
44	34.50	29.43	11.12	1.57	1.17	.49	.31
45	27.67	26.56	11.13	1.35	1.04	.32	.37
46	30.18	26.92	10.27	1.51	1.12	.28	.32
47	26.74	23.56	12.14	.94	1.14	.24	.39
48	27.81	25.03	11.44	1.19	1.11	.29	.22

L2SCA measurement results for SE papers (part 2/2)

Sample	MLS	MLTU	MLC	DC/TU	TU/S	CP/C	NFE/C
49	19.43	18.35	11.20	.65	1.06	.33	.33
50	35.68	35.05	14.80	1.37	1.02	.44	.41
51	31.57	24.56	13.81	.75	1.29	.43	.68
52	26.77	23.58	12.01	.88	1.14	.31	.27
53	27.12	24.75	13.11	.81	1.10	.50	.36
54	22.44	21.02	13.05	.62	1.07	.33	.31
55	25.59	23.76	12.71	.87	1.08	.20	.40
56	23.24	19.55	9.68	1.00	1.19	.24	.24
57	23.20	20.36	11.88	.72	1.14	.50	.39
58	26.91	25.53	14.86	.67	1.05	.54	.38
Mean	26.79	24.14	12.68	.891	1.111	.397	.358
STDev	4.37	3.94	1.79	.269	.061	.130	.091
STDerr.	.57	.52	.23	.04	.01	.02	.01
Median	26.68	23.57	12.37	.84	1.10	.35	.36
Min	18.64	18.14	9.68	.39	1.00	.19	.11
Max	37.68	35.05	18.14	1.57	1.29	.79	.68
Range	19.03	16.92	8.46	1.19	.29	.59	.57

Table 20: Differences between SE and PS measurement

Difference (Δ) = Measure SE paper – Measure PS paper (part 1/2)							
Sample	**ΔMLS**	**ΔMLTU**	**ΔMLC**	**ΔDC/TU**	**ΔTU_S**	**ΔCP/C**	**ΔNFE/C**
1	.959	2.444	.192	.1339	-.082	.024	.0114
2	-1.417	-.877	-.352	.0235	-.018	-.026	-.0943
3	7.292	5.611	3.892	-.0978	.042	.261	.2426
4	1.088	1.219	2.833	-.2163	-.010	-.053	-.3889
5	5.680	6.384	4.773	-.1736	-.027	.272	.1953
6	9.538	8.197	2.384	.2537	.003	-.019	.1650
7	.360	.035	1.203	-.1529	.014	.096	-.0445
8	-6.172	-3.745	-.352	-.3524	-.081	-.023	-.0100
9	3.495	2.000	.614	.0683	.055	.036	-.0818
10	4.739	6.272	2.102	.1528	-.103	.132	.0214
11	-1.730	.459	-.671	.0925	-.105	-.036	.2069
12	-.651	-2.226	.777	-.2427	.077	-.091	-.0161
13	4.996	6.742	2.636	-.0121	-.114	.483	.0963
14	7.134	5.261	.205	.4715	.060	.188	-.0436
15	7.664	4.912	3.151	-.0915	.080	.420	-.0620
16	1.587	3.486	1.118	.1247	-.095	.085	-.1538
17	.221	-.298	-.914	-.0207	.024	-.001	-.1608
18	3.801	3.035	1.479	.1199	.004	.036	-.1076
19	.425	1.272	-.842	.1909	-.039	.004	-.0431
20	3.437	2.954	2.433	-.0607	.012	.044	-.0341
21	-1.746	-1.073	.509	-.0751	-.021	.071	-.0602
22	.238	1.289	-.280	.1545	-.065	-.006	-.0831
23	-1.435	-.123	.548	-.1336	-.063	.144	-.1175
24	1.960	.656	-1.413	.2999	.055	-.002	-.1087
25	1.622	-.171	.309	-.1142	.075	.067	.0593
26	-.870	-.340	-.460	-.0342	-.022	.013	.0911
27	3.430	3.136	-.921	.4010	-.003	.019	.1454
28	-6.060	-6.114	-2.039	-.1618	.015	.029	-.0737
29	-7.717	-4.236	1.233	-.4003	-.097	-.044	-.1028
30	-11.229	-9.597	-1.007	-.5627	-.036	-.235	.0480
31	2.202	3.524	.937	.2040	-.095	.132	-.0264
32	-.197	-.161	.868	-.1270	-.001	.040	.1541
33	-.199	2.303	1.599	-.0135	-.154	.093	.0625
34	-1.091	-1.667	-.430	-.1079	.031	-.015	.0403
35	-2.615	-2.210	1.656	-.3391	-.015	.106	.1376
36	-.719	-3.552	-2.107	-.0046	.146	-.140	-.1034
37	3.937	5.710	.768	.3223	-.145	-.012	.0588
38	2.971	2.341	1.726	-.1059	.013	-.008	-.0825
39	8.408	9.448	3.114	.2247	-.055	.307	.2001
40	1.522	2.715	-.214	.2628	-.081	-.013	-.0566
41	2.449	3.926	-.711	.5281	-.101	-.070	.1343
42	2.599	1.779	-.180	.2468	.024	-.033	.0680
43	-2.352	-2.879	1.693	-.4798	.041	.234	-.0093
44	12.621	10.644	.801	.7527	.008	.080	-.0779
45	-1.275	-.427	-.550	.0494	-.031	.096	.0000
46	5.639	4.328	-1.223	.5703	.035	-.092	-.1839
47	2.669	2.364	1.985	-.0812	-.001	.083	.0636
48	1.569	-.539	1.617	-.3638	.085	.144	.0989

Difference (Δ) = Measure SE paper – Measure PS paper (part 2/2)							
Sample	**ΔMLS**	**ΔMLTU**	**ΔMLC**	**ΔDC/TU**	**ΔTU_S**	**ΔCP/C**	**ΔNFE/C**
49	-1.022	-.725	.052	-.0827	-.013	.041	.0356
50	.661	1.222	.644	-.0045	-.017	.097	.0318
51	.181	-.557	-1.180	.1156	.036	-.047	.2029
52	-.563	-1.975	.079	-.2099	.066	-.024	-.0909
53	-1.605	1.091	.775	-.0816	-.118	.092	-.0959
54	-1.513	-.561	.543	-.0822	-.042	.034	-.1068
55	6.144	5.744	.361	.4195	-.002	-.060	.1874
56	-3.074	-1.743	.554	-.2857	-.047	.021	.0412
57	-5.014	-3.208	2.382	-.6873	-.058	.292	.0362
58	-5.111	-4.990	.908	-.5208	.005	.130	-.0551
Mean	.998	1.181	.683	-.005	-.016	.058	.003
STDev	4.283	3.829	1.435	.285	.063	.127	.119
STDerr.	.56	.50	.19	.04	.01	.02	.02
Median	.543	1.155	.584	-.027	-.012	.035	-.010
Min	-11.23	-9.60	-2.11	-.687	-.154	-.235	-.389
Max	12.62	10.64	4.77	.753	.146	.483	.243
Range	23.85	20.24	6.88	1.440	.300	.719	.631

10 Confirmation of paired sample t-test assumptions

1) Normal distribution for differences between sample pairs (Δ = SE values – PS values)

H_0: Differences between samples are normally distributed ($p > .05$)
H_1: Differences between samples are not normally distributed ($p < .05$)

Tested with Kolmogorov-Smirnov test:

Table 21: Kolmogorov-Smirnov results for paired sample test values

		Differences (Δ = SE paper – PS paper) for each measure							
		ΔMLS	**ΔMLTU**	**ΔMLC**	**ΔDC/TU**	**ΔTU/S**	**ΔCP/C**	**ΔNFE/C**	**N**
Statistics	**Mean**	.997	1.181	.683	-.0052	-.016	.058	.0021	58
	STDev	4.283	3.829	1.436	.2844	.063	.129	.1194	
KS-test Values	**D**	.106	.065	.084	.079	.082	.142	.088	
	p	.531	.951	.807	.863	.825	.192	.764	

H_0 accepted → the distributions of each of the tested samples can be assumed normally distributed.

11 Confirmation of independent sample t-test assumptions

1) Normal distribution of sample values (PS measures)

H_0: Sample is normally distributed ($p > .05$)
H_1: Sample is not normally distributed ($p < .05$)

Tested with Kolmogorov-Smirnov test:

Table 22: Kolmogorov-Smirnov test results for AHS group values

		AHS PS paper values for each measure							
		MLS	**MLTU**	**MLC**	**DC/TU**	**TU/S**	**CP/C**	**NFE/C**	**N**
Statistics	**Mean**	26.44	23.54	12.19	.907	1.128	.340	.355	35
	STDev	4.23	4.02	1.58	.241	.084	.092	.126	
KS-test Values	**D**	.106	.119	.099	.091	.161	.095	.114	
	p	.791	.704	.851	.932	.323	.909	.749	

Table 23: Kolmogorov-Smirnov test results for BHS group values

		BHS PS paper values for each measure							
		MLS	**MLTU**	**MLC**	**DC/TU**	**TU/S**	**CP/C**	**NFE/C**	**N**
Statistics	**Mean**	25.10	22.36	11.62	.907	1.126	.326	.352	19
	STDev	4.19	4.08	1.29	.339	.062	.085	.089	
KS-test Values	**D**	.230	.279	.104	.159	.200	.149	.158	
	p	.230	.103	.973	.722	.432	.794	.730	

H_0 accepted → the distributions of each of the tested measures in each group can be assumed normally distributed.

2) Variances of samples are equal

H_0: Variance of AHS sample is equal to that of the BHS sample ($p > .05$)
H_1: Variance of AHS sample is larger than that of the BHS sample ($p < .05$)

Tested with F-test:

Table 24: F-test results for equality of variances between AHS and BHS samples

		MLS	**MLTU**	**MLC**	**DC/TU**	**TU/S**	**CP/C**	**NFE/C**	**N**
Variance	**AHS**	17.89	16.19	2.50	.058	.0070	.0084	.0159	35
	BHS	17.54	16.64	1.68	.115	.0038	.0072	.0079	19
F-test Values	**F**	1.020	.973	1.489	.507	1.863	1.149	2.006	
	p	.995	.914	.373	.086	.163	.773	.119	
	num df	34	34	34	34	34	34	34	
	denom df	18	18	18	18	18	18	18	
	ratio of Var	1.020	.973	1.489	.507	1.863	1.149	2.006	

H_0 accepted → the variances of the distributions for all measures can be assumed equal between both groups.

12 Confirmation of ANOVA assumptions

1) Normal distribution of sample values (PS measures)

H_0: Residuals are normally distributed ($p > .05$)
H_1: Residuals are not normally distributed ($p < .05$)

Tested with Kolmogorov-Smirnov test:

Table 25: Kolmogorov-Smirnov test results for ANOVA residuals

		MLC residuals	CP/C residuals
Statistics	**Mean**	-1.68E-16	-1.10E-17
	STDev	1.425	.124
KS-test Values	**D**	.071	.149
	p	.914	.136

H_0 accepted → the distributions of each of the tested residuals can be assumed normally distributed.

2) Variances of samples are homogeneous

H_0: Variances in the four groups across the factor courses are equal ($p > .05$).
H_1: Variances in the four groups across the factor courses are not equal ($p < .05$).

Tested with Levene test:

Table 26: Levene test results for homoscedasticity of ANOVA group variances

		MLC ~ courses	CP/C ~ courses
Levene-test Values	**Df 1**	3	3
	Df 2	54	54
	Test value	1.150	.295
	p	.338	.829

H_0 accepted → the variances of the four groups can be assumed equal.

13 Paired sample t-test for German as L1 samples only

Table 27: Difference of syntactic complexity between PS paper and SE paper for samples with only German as L1 (N=48)

Sub-construct	**Measure**	**PS mean (SD)**	**SE mean (SD)**	***p* (t-value/df)**	**Cohen's *d*** [b]
Overall sentence complexity	MLS	26.11 (4.33)	27.17 (4.42)	.1083 (-1.6371/47)	.2405
Overall T-unit complexity	MLTU	23.35 (4.15)	24.55 (4.00)	.0442 (-2.0676/47)	.2952
Clausal subordination (finite)	TU/S	1.12 (.077)	1.11 (.062)	.1201 (1.5829/47)	-.2118
Elaboration at clause level	DC/TU	.914 (.287)	.903 (.276)	.7935 (.26321/47)	-.0392
Clausal coordination	MLC	12.08 (1.53)	12.83 (1.85)	.0007* (-3.6192/47)	.4429
Phrasal coordination	CP/C	.336 (.091)	.397 (.131)	.0034* (-3.0882/47)	.5321
Non-finite elements/subordination	NFE/C	.366 (.112)	.364 (.086)	.9399 (.07579/47)	-.0125

* significant result for $p \leq .007$

[b] $d < 0$ indicates decline and $d > 0$ indicates rise in means from PS to SE paper for these values.

14 One-way ANOVA result tables

Table 28: One-way ANOVA results for ΔMLC with factor no. of courses

	df	**Sum of squares**	**Mean square**	**F**	**Sig.**	**η^2**
no. of courses (btw. groups)	3	1.615	.538	.251	.860	.014
residuals (within groups)	54	115.766	2.144			
Total	57	117.382				

Table 29: One-way ANOVA results for ΔCP/C with factor no. of courses

	df	**Sum of squares**	**Mean square**	**F**	**Sig.**	**η^2**
no. of courses (btw. groups)	3	.049	.0163	1.006	.397	.053
residuals (within groups)	54	.877	.0162			
Total	57	.926				

15 Abstract (English)

This diploma thesis is concerned with measuring syntactic complexity in academic papers of English students at the University of Vienna. It investigates if syntactic complexity changes over the course of their studies and if EAP instruction contributes to this change. Also, the influence of factors such as writing instruction and previous language education and genre exposure on syntactic complexity is investigated. Syntactic complexity has been described as one sub-dimension of the multicomponential framework of complexity, accuracy and fluency (CAF). Moreover, syntactic complexity is seen as a useful indicator for second language (L2) writing development. In this thesis, syntactic complexity is operationalized as a multi-dimensional construct based on objective measures that is able to provide information about certain aspects of syntactic development at the sentence, clausal and phrasal level.

The investigated data set comprises 58 samples which each consists of two different academic papers and additional learner background information. All samples have been analyzed with the help of the L2SCA analysis software (Lu 2016), which has been validated in the course of the analysis. The findings support the pattern of syntactic development as suggested by previous research that syntactic complexity develops mainly at the clausal and phrasal level in highly proficient L2 writers. Furthermore, the results imply valuable information for further research about the influence of previous text genre experience and Austrian secondary school types on the syntactic development in the L2.

16 Abstract (German)

Diese Diplomarbeit behandelt das Messen von syntaktischer Komplexität in Seminararbeiten von Englischstudentinnen und Englischstudenten an der Universität Wien. Dabei wurde die Veränderung syntaktischer Komplexität zwischen zwei Seminararbeiten analysiert und ob akademische Schreibkurse zu dieser Veränderung beitrugen. Weiters wurde untersucht welchen Einfluss die Art der oberen Sekundarstufe (AHS oder BHS) auf die weitere Entwicklung von syntaktischer Komplexität in den Probandinnen und Probanden hatte. Syntaktische Komplexität wird als eine Dimension der drei Komponenten Komplexität, Genaugkeit und Flüssigkeit (CAF) gesehen. In diesem Sinn ist syntaktische Komplexität ein Indikator für die schriftliche Entwicklung der Zweitsprache Englisch. Diese Diplomarbeit verwendet mehrere objektive Messparameter um syntaktische Komplexität auf Satz-, Gliedsatz- und Phrasenebene abzubilden.

Das untersuchte Korpus umfasst 58 Datensätze, die je zwei Seminararbeiten und zusätzlichen Informationen über die Probandin oder den Probanden enthalten. Die Daten wurden mithilfe der Analysesoftware L2SCA (Lu 2016) analysiert. Im Zuge der Analyse wurde auch die Analysesoftware validiert. Die Ergebnisse untermauern die bisherigen Forschungsergebnisse, dass Veränderungen in syntaktischer Komplexität hauptsächlich auf der Gliedsatz- und Phrasenebene im untersuchten Sprachniveau ausgeprägt sind. Weiters liefern die Ergebnisse wertvolle Informationen für weitere Forschung über die Auswirkungen des österreichischen Sekundärschultyps in Bezug auf syntaktische Komplexität akademischer Texte in der Zweitsprache.

Table 16: Sumary of L2SCA and Manual annotations for annotated samples

L2SCA annotations vs. manual annotation																				
Sample	**TU**	**T_M**	**Delta**	**ident.**	**C**	**C_M**	**Delta**	**ident.**	**DC**	**DC_M**	**Delta**	**ident.**	**VP**	**VP_M**	**Delta**	**ident.**	**CP**	**CP_M**	**Delta**	**ident.**
1_PS	96	98	2	96	153	155	2	153	53	57	4	53	208	208	0	208	51	52	1	51
1_SE	86	87	2	86	151	149	2	149	59	62	4	59	207	204	0	204	54	56	1	54
30_PS	60	62	1	60	117	117	2	117	53	54	3	53	166	166	3	166	51	51	2	51
30_SE	78	83	2	78	143	143	0	143	58	58	1	58	191	190	0	190	66	66	0	66
60_PS	85	87	5	85	211	210	0	210	120	120	0	120	285	282	1	282	44	43	0	43
60_SE	98	102	2	98	168	171	1	168	71	69	0	69	233	231	3	231	84	88	1	84
Sum	503	519	14	503	943	945	7	940	414	420	12	412	1290	1281	7	1281	350	356	5	349
Mean	83.8	86.5			157.2	157.5			69.0	70.0			215.0	213.5			58.3	59.3		
STDev	13.8	14.0			31.3	31.2			25.8	25.0			40.8	39.8			14.5	15.9		
STDerr	5.65	5.73			12.76	12.74			10.55	10.22			16.65	16.26			5.91	6.50		
Median	85.5	87.0			152.0	152.0			58.5	60.0			207.5	206.0			52.5	54.0		
Min	60	62			117	117			53	54			166	166			44	43		
Max	98	102			211	210			120	120			285	282			84	88		

Table 17 Correlation coefficient between measures calculated from L2SCA and manual annotations

Correlation results between measures calculated from L2SCA and manual annotations														
Paper	**MLS**	**MLS_M**	**MLTU**	**MLTU_M**	**MLC**	**MLC_M**	**DC/TU**	**DC/TU_M**	**TU/S**	**TU/S_M**	**CP/C**	**CP/C_M**	**NFE/C**	**NFE/C_M**
1_PS	24.3	24.3	20.7	20.3	13.0	12.8	0.552	0.582	1.171	1.195	0.333	0.531	0.359	0.342
1_SE	25.2	25.2	23.2	22.9	13.2	13.4	0.686	0.713	1.089	1.101	0.358	0.644	0.371	0.369
30_PS	33.0	33.0	31.3	30.3	16.1	16.1	0.883	0.871	1.053	1.088	0.436	0.823	0.419	0.419
30_SE	26.9	26.9	25.5	24.0	13.9	13.9	0.744	0.699	1.054	1.122	0.462	0.795	0.336	0.329
60_PS	28.2	28.2	23.6	23.0	9.5	9.5	1.412	1.379	1.197	1.225	0.209	0.494	0.351	0.343
60_SE	23.2	23.2	20.4	19.6	11.9	11.7	0.725	0.676	1.140	1.186	0.500	0.863	0.387	0.351
Mean	26.8	26.8	24.1	23.3	12.9	12.9	0.8	0.8	1.1	1.2	0.4	0.7	0.4	0.4
STDev	3.5	3.5	4.0	3.8	2.2	2.2	0.3	0.3	0.1	0.1	0.1	0.2	0.0	0.0
STDerr	1.44	1.44	1.64	1.56	0.89	0.90	0.12	0.12	0.02	0.02	0.04	0.06	0.01	0.01
Median	26.1	26.1	23.4	23.0	13.1	13.1	0.7	0.7	1.1	1.2	0.4	0.7	0.4	0.3
Min	23.2	23.2	20.4	19.6	9.5	9.5	0.6	0.6	1.1	1.1	0.2	0.5	0.3	0.3
Max	33.0	33.0	31.3	30.3	16.1	16.1	1.4	1.4	1.2	1.2	0.5	0.9	0.4	0.4
Covariance	12.39		15.25		4.78		0.0871		0.0033		0.0157		0.0009	
Pearson's *r*	1.000		0.994		0.998		0.994		0.950		0.942		0.910	
p	<2.2E-16***		0.00003918***		0.000007679***		0.00005452***		0.004**		0.005**		0.012*	
t (df =4)	1.34E+08		19.698		29.675		18.122		5.9531		5.6483		4.4197	

Significance level: ***$p \leq .001$, **$p \leq .01$, *$p \leq .05$