Creating a Classroom-MT: Connecting Simplification Methods to Language Learner Levels in Monolingual Machine Translation

Utrecht University
Faculty of Humanities
Bachelor of Science - Artificial Intelligence
Naomi Langstraat - 4147804
Supervisor: Dr. Rick Nouwen - *Utrecht University*
2nd Reader: Dr.habil Cassio P. de Campos - *Utrecht University*
July 2019

Abstract

This study aims to investigate whether there is a theoretical basis for the creation of a Classroom-MT, that would be able to simplify high level texts to the level of an L2 language learner. This paper will classify language learners based on the CEFR levels presented by the Council of Europe (2001; 2018) and variables that determine the developmental sequence of an L2 presented by Goldschneider and DeKeyser (2001). It will then give an overview of state-of-the-art (monolingual) machine translation models. The models will be examined both amongst each other as well as their usefulness in the creation of the Classroom-MT. This study concludes that there is a stable theoretical basis to create a Classroom-MT, in both the language learner levels as well as the simplification techniques available. Future research is needed in both the search for quantifiable variables connected to the developmental sequences, as well as how this Classroom-MT can be implemented correctly.

Keywords: CEFR, monolingual machine translation, text simplification, developmental sequences, L2 learner levels, machine translation, phrase-based model, tree-based model

Contents

1 Introduction

1.1 Theoretical Background

Machine translation has been used as a tool in foreign language education for years. It is not new, but the tool has been used for the ´wrong´ reasons. The use of machine translation as a so-called bad-model is well documented in the literature (Niño, 2009; Garcia & Pena, 2011, and many more). The bad-model approach uses a machine translator (MT) as an error-ridden text generator. Students are given a text generated by an MT and are asked to find mistakes and change the text according to what they have learned about the language so far. Although the bad-model use of machine translation has been found to be beneficial in language learning, there are two problems with this approach. It exposes foreign language learners with faulty and flawed texts (Niño, 2009), and it does not take into account the fact that machine translation has seen a big improvement in the last 10 years (Statt, 2016; Gershgorn, 2016; Quoc & Schuster, 2016).

A positive way in which MTs have been used in recent times is in the creation of monolingual machine translators (MMT). MMTs make use of simplifications methods to create an easier version of a text in a certain language. Simplification methods are not new and have been used for generating easier to read texts for aphasic or deaf patients, as well as people with low reading abilities, and learners of a second language (Quigley & Paul, 1984; Caplan, 1992; Parr, 1993; all as cited in Siddharthan, 2006). A relatively well-known example of simplified language is Simplified English, used in Simple Wikipedia pages.[1] Wikipedia believes Simple Wikipedia is used, amongst others, by second language learners of English (Simple English Wikipedia, 2019). As said before, recent times have seen a rise of MMTs that try to create these simplified versions automatically based on simplification rules. This research will try to connect literature on proficiency levels of language learners with the

[1] An example of Simplified English can be found https://simple.wikipedia.org/wiki/Star_Wars (simplified version) and can be compared to https://en.wikipedia.org/wiki/Star_Wars (“normal” version).

literature on simplification techniques and MMT. This is done to answer the following research question: *How can we create an MMT that, when given the reading comprehension level of the language learner and a certain text, produces a simplification of the given text that is in accordance with the given level?* The following sub-questions have been formulated to answer this question. These will be discussed separately in this paper.

1. How can we classify L2 language learners in terms of reading comprehension?
2. What simplification methods and language models are used in monolingual machine translation?
3. How can we connect learner levels to simplifications in monolingual machine translation?

Thus, this study aims to do three things. Firstly it will aim to find so-called developmental sequences and use it to classify language learners reading comprehension. This part will focus on developmental sequences in foreign language learning and introduce the so-called CEFR levels. Secondly this paper aims to provide insights in how simplification methods work and how they can be used. This part will introduce several of the well-known simplification methods found in the literature and explain why they may be useful in the task of MMT. And thirdly this paper aims to give advice on the creation of a *Classroom-MT* and give ideas on techniques used to simplify language that will be beneficial in the creation of a *Classroom-MT* specifically.

This *Classroom-MT* will allow teachers to input a high-level text and the MT will return a simplified version based on the level of the class or individual learner. This will range from extreme changes to the text for a beginning learner and more minor changes for an intermediate learner. The advice given in this paper will add to the existing literature in two separate ways. Firstly, it will connect both lexical and syntactical methods of simplification where existing literature has either focused on syntactical or lexical methods. Secondly, it will connect simplification methods to existing literature in the field of developmental sequences, giving an opening into the creation of MMT that will give simplifications on different levels. Existing literature in this field has had two levels, the original level and the simplified level.

1.2 Relevance to Artificial Intelligence

Machine translation is one of the many ways in which AI is trying to help human specialists in taking over certain tasks that were deemed profoundly human-like. In recent years machine translation has, as cited above, gained a lot in terms of human-like behaviour. The place of AI is rapidly growing in our everyday life. Many applications are made to do a certain task. This study aims to use one of these applications, machine translation, which was made for the task of translating, in another task; learning. This paper will try to give an insight in the quality and usability of machine translation in an every-day task such as foreign language learning and teaching.

To be able to create and use these kinds of new applications of AI it is important that one is not only aware of the technological innovations necessary but also the understanding of the domain it is applied to. In the case of this paper it is important to both understand the technological practices used in MMTs, and the theory of reading comprehension levels in second language acquisition.

2 Classifying the language learner

To be able to give the right simplified translations to the right students, we need to do two different things. We, firstly, need a method to classify language learners in categories that are in accordance with their language level. This will be discussed in section 2.1 of this paper. Additionally, we need to be able to explain what lexical and syntactical levels are met in these levels. This will be discussed in section 2.2 of this paper.

2.1 How to classify the language learner in levels

Classifying language learners has been done extensively for a few decades now. Best known are probably the Cambridge English Exams for the classification of language learners who wish to learn English. But other languages also have their language-specific test to classify learners; a few examples are DELE (Spanish), CNaVT (Dutch), TCF (French), and TestDaF (German). Almost all of these tests now refer to the level one would receive on the CEFR scale in their results. This section will firstly explain what the CEFR is (2.1.1), and will then explain how classifying learners based on reading comprehension works (2.1.2).

2.1.1 CEFR Levels

In 2001 the Council of Europe launched the so-called *Common European Framework for Reference for Language* (from now on CEFR, its official abbreviation) (Council of Europe, 2001). This framework puts language learners in categories that are in accordance with their language abilities (focused on what the language learner *can* do). For decades researchers have tried to do this for every language separately, as shown above. The fact that this framework works language independently made the Council of Europe call this a “major turning point” (Council of Europe, 2018a). For the first time a language-level reference guide could be applied to all languages. The Council of Europe positions the CEFR as a transparent non-biased and clear basis for testing the proficiency of language learners (Council of Europe, 2018b). It is meant to distinctly describe what a language learner should be able to do, and what knowledge is needed to do so effectively (Council of Europe, 2001). And so in 2001 the CEFR defined 3 categories with each 2 subcategories in which language learners can be placed, based on their knowledge.

The Council of Europe explains the different levels in the following manner (2001). The CEFR levels, range from A to C where all letters have two sub-categories. This results in the six categories named; A1, A2, B1, B2, C1 and C2. These categories are in progressing order; A1 (*Breakthrough*) and A2 (*Waystage*) are levels that are concerned with starting foreign language students; the “Basic User”. B1 (*Threshold*) and B2 (*Vantage*) are concerned with the “Independent User”, the student that is further along in his or her studies and is able to communicate relatively fluently, often called an intermediate learner; and students who are in C1 (*Effective Operational Proficiency*) or C2 (*Mastery*) carry the name “Proficient User”, where C2 corresponds to native-like abilities in a foreign language (see Figure 1). The Council of Europe provides many tables and figures to clarify these 6 levels.

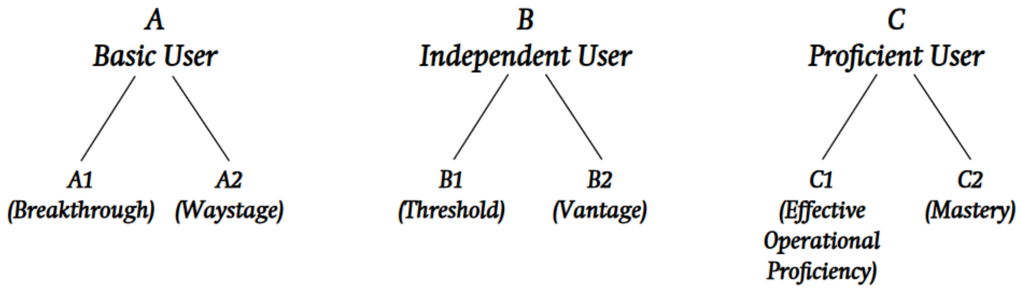

Figure 1 This figure shows the different layers in the CEFR, horizontally it shows the advancement of a language learner from Basic User to Proficient User. Vertically it shows the different categories named inside the three different User-Levels (Council of Europe, 2001 p.23).

These CEFR tables can be found in roughly two different categories. The *Global Table* (Figure 2) gives an overview of what a language learner should be able to do in the different categories, regardless of what language he or she is learning. This is a language-independent reference guide and what makes CEFR special. And *Self-assessment tables* can be found language- and task-specifically (see Figure 3). This means that a student of Spanish can, for instance, access what he or she should be able to do in level B2, and specifically for writing. In this paper we are mainly interested in Reading Comprehension. The CEFR combines both Listening and Reading in the category called *Understanding*. The place and importance of reading comprehension in foreign language education will be discussed in the next section (2.1.2).

PROFICIENT USER	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
INDEPENDENT USER	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
BASIC USER	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

Figure 2 The global table provided by the Council of Europe that is language-neutral and gives an indication of what a language learner should be able to do regardless of the language in question (Councile of Europe, 2018c).

		A1	A2	B1	B2	C1	C2
UNDERSTANDING	**Listening**	I can recognise familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.	I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements.	I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.	I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programmes. I can understand the majority of films in standard dialect.	I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. I can understand television programmes and films without too much effort.	I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided I have some time to get familiar with the accent.
	Reading	I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues.	I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.	I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.	I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.	I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field.	I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.

Figure 3 English specific Understanding category for self-assessment containing the subcategories listening and reading (council of Europe, 2018d).

To give an example of what certain texts would look like, Table 1 gives two starting paragraphs of two different levels of text. A quick analysis gives the statistics presented in Table 2. Although it is only 1 paragraph of the text, and this is just an indication, it shows that the sentences in a B2 text are substantially longer and the words are also longer than the words in the A1 text. Also content-wise the texts differ greatly. As mentioned in A1 level in the Global Table (Figure 2) A1 learners are able to deal with simple introductions, which is exactly what this level A text is about.

Table 1 First paragraph of Level A1 and Level B2 Text (found: see footnotes)

Level A1 Text	**Level B2 Text**
My name is John Hi! Nice to meet you! My name is John Smith. I am 19 and a student in college. I go to college in New York. My favorite courses are Geometry, French, and History. English is my hardest course. My professors are very friendly and smart. It's my second year in college now. I love it![2]	**Boston** Jean and her family recently traveled to Boston, Massachusetts, one of America's oldest colonial cities. Boston is rich in history and local personality. During their visit, Jean and her family appreciated learning about Boston's role during the American Revolution.[3]

Table 2 Statistics of Example A1 and B2 Text

	Sentence Length	**Token Length**
Level A1 Text	5.5	4.15
Level B2 Text	13	5.85

The CEFR is now a commonly used framework in language learning on all different levels (Baldwin & Appelgren, 2018). Harsch and Rupp (2011) call it a "key reference document" for teachers and students alike. Baldwin and Appelgren show that much research has provided us with the successful accomplishments of the CEFR in the making of language assessment tests. It seems the CEFR has taken its place as the prime language assessment and learner classification method. But a widely used reference guide will not bypass its share of criticism. Baldwin and Appelgren (2018) mention that the use of "vague and imprecise language" in the CEFR is one of the biggest problems. Examples of this sort of language are found in Figure 2, where A1 learners are told they should be able to understand "basic phrases", B1 learners can handle "clear standard input", and C1 learners understand "longer"

[2] https://lingua.com/english/reading/john/
[3] https://lingua.com/english/reading/boston/

texts. What are basic phrases? What is a clear standard input? How long is a longer text? This is not explained any further and this sort of vocabulary can come across as questionable and uncertain.

So the problem is that the CEFR levels do not give the precision of language levels that are needed for formalisation. The challenge in this first part of the paper will, therefore, lie in finding the right definitions for these *vaguely and imprecisely* described levels. To be able to create a Classroom-MT that uses these levels it is important to be more clear and more precise because these levels will need to be formalised into programmable variables. This paper will try to do two things. Firstly, it will attempt to give more precise definitions of this imprecise language in terms of what it would mean for vocabulary and length of the text (see section 2.3). Secondly, this paper will seek to give a formalizable definition of so-called developmental sequences (section 2.3) based on the theory found. The variables that are presented will give an indication on how language levels can be manipulated. This is done to give a starting point for the quantification of these variables, it will remain an empirical question what values this variables will have. And what values will result in either a level A or B or C text.

2.1.2 Reading Comprehension

This article will use the following definition of reading comprehension

> "Reading comprehension involves abilities to **recognize words rapidly and efficiently**, develop and use a **very large recognition vocabulary**, **process sentences** in order to build comprehension, engage a range of strategic processes and underlying cognitive skills (e.g., setting goals, changing goals flexibly, monitoring comprehension), interpret meaning in relation to background knowledge, interpret and evaluate texts in line with reader goals and purposes, and process texts fluently over an

> extended period of time. These processes and knowledge resources allow the reader to generate text comprehension to the level required" (Grabe, 2014, p.8) (bold added).

As becomes clear by the definition above, reading comprehension entails many different qualities for a language learner to possess. To emphasise a few of these qualities, some phrases have been made bold in the text. According to Grabe (2014) it is important to recognize words, have a large vocabulary and to be able to process sentences. These three points touch on very important skills in foreign language learning, namely vocabulary development, syntax (processing sentences), and semantics (being able to recognize words in their environment and extract meaning from that). All these skills come together in reading comprehension, and that is why it is an interesting language-skill to investigate. Grabe has shown that research on L2 vocabulary knowledge and L2 syntax have both shown that there is a strong linkage between the degree of knowledge on these subjects and the level of reading comprehension.

In the process of learning to read in a foreign language the goal should always be to become a so-called 'fluent reader' (Grabe). A fluent reader should, according to Grabe, be able to (1) read faster than he or she can listen, (2) have a very large vocabulary knowledge and (3) be able to automatically, and at the same time, link the syntax of a sentence to its meaning. The CEFR reading comprehension levels do not speak of speed (1). Linking the other two (2 and 3) demands by Grabe to the CEFR reading comprehension levels presented in Figure 3 we can conclude that at every level the language student becomes a fluent reader in one or more forms of text. A language student will learn the necessary vocabulary, syntax and semantics needed for these types of text in these specific levels. For A1 this would be poster text, for A2 it is short personal letters, for B1 longer personal letters (among others), and so on.

2.2 Developmental Sequences

It is widely accepted that people learn their first language (L1) in a certain developmental sequence (Tomasello, 2000; Pinker, 2013). The question that arises is whether, in the acquisition of a second language (L2), there is also such a thing as a developmental sequence. And if it is present in what way this is linked to (1) the L1 developmental sequences that we know and (2) the L1 that the learner already possesses. If it is the case that L2 developmental sequences exist, and this is the reason this paper covers this, this would be interesting for the Classroom-MT. In the creation of this MT it is important to be able to differentiate between the different output-levels we would want to create. For instance, how a level A text differs from a level B text. Developmental sequences in second language learning would give a method to formalize how a level A learner differs from a level B learner. Suppose Anne, a level A learner, knows Z but does not know W, and suppose Mark, a level B learner, knows both Z and W. We would want our Classroom-MT to create a level A text without W (potentially with Z), but a level B text can include W. If there are developmental sequences in L2, the Classroom-MT would be able to leave *out* or leave *in* certain formalizable features of a text.

Research shows that just as in L1 acquisition, L2 acquisition has developmental sequences. This means that “what is learned early by one is learned early by others” (Lightbown & Spada, 2013). Some examples of these sequences will be discussed later in this section. But since developmental sequences have been found the two questions raised above will have to be answered. Firstly, the question how the L2 developmental sequences are linked to the well-known L1 developmental sequences. It is important to note, and make very clear, that while L2 developmental sequences exist they are *not* the same as L1 developmental sequences. It means that regardless of the L2 that is learned, a student will go through the same developmental sequences in learning a second language (Lightbown & Spada, 2013).

The second question raised above, may be more difficult to answer. It is widely accepted that whilst learning an L2 students will experience L1 transfer. A form of L1 transfer is found in the Alternation Hypothesis which states that, if language learners can choose between two patterns in an L2, they will choose the pattern that also exists in their L1 (if that is the case) (Jansen, Lalleman & Muysken, 1981). Jansen, Lalleman and Muysken show that this is only partly true, because there are differences in the use of a feature in a second language that can be linked to L1 transfer, but their research also showed that this language transfer does not influence the general (L1 independent) preference of an L2.

They researched the difference in word order between Turkish and Moroccan learners of Dutch. In a Dutch main clause with simple verbs the word order is SVO and in main clauses with auxiliary verbs the order is SOV. The main word order in Turkish is SOV and Moroccan Arabic has VSO and SVO as the main word orders. Jansen et al. show that Moroccan learners use more SVO word order than Turkish learners of Dutch. This is accounted for by their L1 word order. However, both groups still prefer to use SOV word order in Dutch. Zobl (1982) adds that English learners of Dutch portray the same preference for SOV word order in Dutch. This is interesting for this research, because in the search of L2 developmental sequences, it shows people still follow certain patterns in L2 acquisition, even though their L1 does influence L2 usage.

According to Zobl (1982), L2 learners will go through two different stages when it comes to grammatical morphology regardless of the L1 they possess. Firstly they will go through a period in which a certain grammatical morphology is not present at all in their utterances. This can be, for instance, inflections or articles. Zobl argues that when a language has something similar, this stage is shorter for a learner with that L1. The next stage is the stage in which the elements occur but they do not always occur and not always correctly. Zobl states that if there is nothing similar in a language, then this stage too may be longer for a learner from that L1 compared to a learner with an L1 that has something similar.

These researchers show that L1 can indeed have an influence but, and this is backed up with more recent research (cited in Lightbown & Spada, 2013), research in this area shows that L1 transfer does not interfere with the order of acquisition but more in the speed of acquisition and/or starting sequence.

This speed of acquisition, or a different starting sequence, does not have an influence on determining the reading level based on the developmental sequences. It means that students with a certain L1 will take longer to get to a higher reading level on the CEFR scale compared to another student whose L1 may be more closely related to the L2. The above only shows that the speed of going up the CEFR scales is affected by the L1 background.

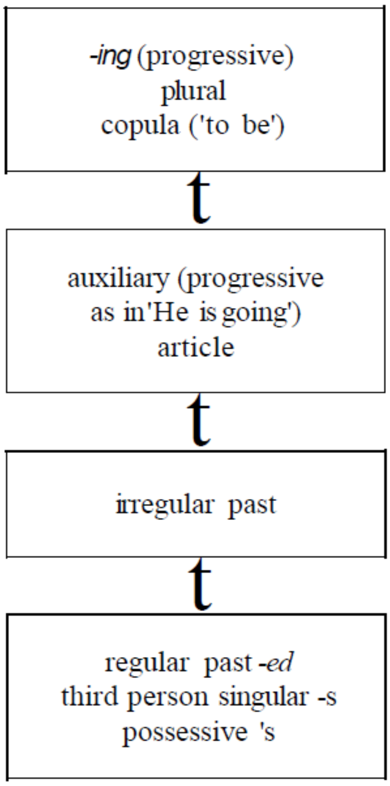

Figure 4 Developmental sequence of acquisition of morphemes for L2

These developmental sequences can be found in all different parts of language acquisition. An example is the patterns in the development of syntax and morphology. They were not the same as the patterns for L1, but similar among L2 learners (Lightbown & Spada). Figure 4 shows the developmental stages of the acquisition of morphemes for L2 learners of English. One can see that the progressive '-ing', the plural, and the copula 'to be' are learned first, and the regular past '-ed', the third person singular '-s' and the possessive ''s' are learned last.

Researchers have also found developmental sequences in other parts of language such as negation (Schumanm, 1979; Wode, 1978), questions (Pienemann, Johnston & Brindley, 1988), possessive determiners (White, 1998 & 2008)), relative clauses (Gass, 1982), and reference to the past (Meisel, 1987; Collins, 2002).

An example of the developmental sequences in question forming is given by Pienemann and colleagues as they present 6 different stages that learners will go through. They give the following stages: (1) single words; "Dog?", (2) Declarative word order; "The boys throw the shoes?", (3) Fronting; "Where the children are playing?", (4) Inversion in wh- + copula; "Where is the sun?", (5) Inversion in wh- questions; "How do you say 'proche'?", and (6) Complex questions such as question tags, negative questions and embedded questions; "It's better, isn't it?".

Another example would be the developmental sequence of relative clauses. Table 3 shows the different parts of speech and their corresponding relative clauses. Gass (1982) observed that if learners of a language could use, for instance, the relative clause connected to the indirect object, they would be able to use any relative clause above that one. She also found that it worked the other way, if they were not able to use, for instance, the relative clause connected to the direct object, they were not able to use any relative clauses further down the list.

Table 3 "Accessibility hierarchy for relative clauses in English" (Lightbown & Spada, p.54)

Part of Speech	**Relative Clause**
Subject	The girl who was sick went home.
Direct Object	The story that I read was long.
Indirect Object	The man who[m] Susan gave the present to was happy
Object of Preposition	I found the book that John was talking about.
Possessive	I know the woman whose father is visiting.
Object of Comparison	The person that Susan is taller than is Mary

From the papers, studies, analyses, and many examples coming from them, we can conclude that developmental sequences are the norm, not only in L1 acquisition, but also in L2 acquisition. This is interesting because this means that a machine translator can be programmed to make use of different levels of learning in L2. To be able to use these developmental sequences in our Classroom-MT, we will need to investigate how these developmental sequences can be formalised. It is, therefore, interesting to find what this developmental order is based on. What makes some parts of language more difficult than others? How can we language independently say something about the difficulty of a language phenomenon? If it is possible to find certain patterns that make up this developmental order, it might be possible to formalize these patterns of language. Formalized patterns would be easier to program into a machine translator.

Goldschneider & DeKeyser (2001) identify a number of language variables that contribute to L2 developmental order (see Table 4). This means that of the examples mentioned above (developmental order of questions, relative clauses etc.) the reason for the order is likely to be a result of one or more of the variables mentioned here. *Perceptual Salience* (1) refers to how striking a feature is in a language; is it easy to spot? *Semantic complexity* (2) is the degree of difficulty of a feature on the basis of how many meanings it conveys. *Morphophonological Regularity* (3) is based on the fact that a phonological environment may influence the phonological representation of a feature. *Syntactic Category* (4) scores a certain language feature on the basis of its place in the so-called Functional

Category Theory. This theory described the order of acquisition on the basis of whether the feature is lexical (first) or functional (second), and within the categories whether the feature was free (first) or bound (second) (Goldschneider & De Keyser, 2001). And lastly, *Frequency*(5) refers to how often a certain feature occurs in an L2.

Table 4 Language variables that may contribute to developmental order of L2 acquisition

1.	Perceptual Salience
2.	Semantic Complexity
3.	Morphophonological Regularity
4.	Syntactic Category
5.	Frequency

To be able to clearly define what is needed for a machine translator and how Goldschneider and DeKeyser formalised these language features, the first four variables will be described in more detail. Frequency is the most straight-forward of the 5 variables and does not need further explanation. Goldschneider and DeKeyser came up with a scoring system for each variable to be able to give a complexity score to a word, sentence and/or text. For frequency this would mean that a frequent word would receive a high score, and a less frequent word a lower score. They based this on earlier research done by Brown (1973; as cited in Goldschneider & DeKeyser), who has made a complete scoring system for frequency.

Goldschneider and DeKeyser characterise Perceptual Salience as how easy it would be for a learner to either hear or see the part under investigation (from now on PUI). They state the higher the degree of perceptually salience as PUI is, the earlier it is acquired. They base the score of perceptual salience on the following variables: (1) number of phones, (2) presence/absence of a vowel in the surface form (syllabicity), and (3) total relative sonority. Points were given for all of these variables. Because Perceptual Salience is considered vague by many authors, Goldschneider and DeKeyser give the following examples:

(1) Number of Phones: The plural -s has three different pronunciations, [s], [z] and [əz]. Thus the plural -s has three allomorphs, but 4 phones; [s], [z], [ə], and [z]. The number of phones is calculated by dividing the number of actual phones by the number of allomorphs.

(2) Syllabicity: Again, the plural -s has three allomorphs of which only one contains a vowel; [əz]. Thus the score for syllabicity, for the plural -s, is 0.33 (number of allomorphs divided by allomorphs with a vowel in the surface form).

(3) Sonority: These scores are based on the sonority hierarchy of Laver (1994) seen in Table 5.. The more sonorous the more points are awarded. Examples: Mid Vowels receive 8 points, Fricatives 3 points. The complete PUI is given points and they are added up.

Table 5 Sonority Hierarchy Laver (1994, p. 504)

Most sonorous		
	Vowels	Low Vowels
		Mid Vowels
		High Vowels
	Glides	
	Liquids	
	Nasals	
	Obstruents	Fricatives
		Affricates
		Stops
Least Sonorous		

Semantic complexity is defined as “how many meanings are expressed by a particular form” by Goldschneider and DeKeyser (p. 24). They specify that the more complex a PUI is the longer it takes for a language learner to acquire. Points are given based on the assumption that the more meanings a PUI has the more difficult it is to learn compared to PUIs with fewer meanings. A point is awarded for every meaning a PUI has. The plural -s receives 1 point from Goldschneider and DeKeyser because it only expresses number. They give 3 points to the 3rd person singular -s because it expresses person, number, and tense.

Morphophonological regularity is characterised, by Goldschneider and DeKeyser. as the extent to which the phonological environment affects the PUIs. Goldschneider and DeKeyser explain that being phonologically regular means a PUI will receive a low score and is earlier required than PUIs with higher scores and lower regularity. Points are awarded based on two characteristics which are *phonological alterations* and *homophony*. Phonological alterations are the different ways in which a PUI can be pronounced based on their phonological environment. The PUI *the* will receive a score of 2 because it can either be pronounced as [ðə], as in *the car*, or as [ði], as in *the apple* (Goldschneider & DeKeyser). Goldschneider and DeKeyser award 1 point per alternation. They assign either 1 (not homophonous) or 2 points (homophonous) based on the homophony of a PUI. Homophony is whether or not a PUI has a counterpart that sounds exactly the same but means something different. According to their scoring system the plural -s would get 2 points because it is homophonous with the third singular -s and the possessive ‘s. The article *the*, mentioned above, would get 1 points because it is not homophonous.

Table 6 Points assigned to syntactic catgegories (Goldschneider & DeKeyser, 2001)

Lexical	
Free	4 points
Bound	3 points
Functional	
Free	2 points
Bound	1 point

Goldschneider and DeKeyser give the numbers in Table 6 for classifying syntactic categories. This means that a category such as a plural -s will receive 3 points for being a bound lexical item. A third singular -s will receive only one point for being a bound functional item. We can see this back in Figure 4 where the plural is learned first and the third singular -s is learned last.

Even though it is remarkable that just as in L1 acquisition, there is a developmental order in L2 acquisition, Lightbown and Spada (2013) emphasize that “developmental stages are not like closed rooms”. They argue it is perhaps better to think of a stage as being “characterized by the emergence and increasing frequency of new forms rather than by the complete disappearance of earlier ones”. Note, for instance, that when a learner progresses to a higher stage in their learning process, it does not mean that they will always produce fewer errors (2013). Lightbown and Spada continue that even when learners move to “a more advanced stage” and different features become dominant, certain conditions such as stress or complexity may have the effect that a learner falls back into old patterns learned in an earlier stage.

2.3 Connecting Developmental Variables to Language Levels

The three levels that this paper will cover, and therefore the three levels that the Classroom-MT should be able to create, will be named Level A (corresponding to A1 and A2), Level B (corresponding to B1 and B2) and Level C (corresponding to C1 and C2). The Classroom-MT should be used to simplify any Level C text to either a Level A text or a Level B text. This chapter will try to define all levels as detailed as possible. The second part of this paper will be used to dive into the different techniques that can be used to simplify texts to these levels.

In addition to these 3 levels we now have 5 components that determine developmental sequences (Table 4). This goal of this section is to give a clear definition of the CEFR levels as well as what that means for the components mentioned by Goldschneider and DeKeyser (2001). This means that two things will have to be done. Firstly the CEFR levels give an idea of what a text should look like at different levels. Using the guidelines given in Figure 3 definitions of these levels will be made. And secondly this section will then give explanations of what the components given in Table 4 (given again for clarity) will mean in these levels.

Table 4 (copy) Language variables that may contribute to developmental order of L2 acquisition

1. Perceptual Salience
2. Semantic Complexity
3. Morphophonological Regularity
4. Syntactic Category
5. Frequency

The idea behind this section is to give a method to determine the value of a text (Level A, B, or C). This should be done on the basis of scores given to the developmental components. As Goldschneider and DeKeyser (2001) showed it is possible to give scores to the variables listed in Table 4. This section will give logical formulas that will explain what a Level X text should adhere to.

There are three ways to approach the scores given; (1) We could choose to look at the average word score, (2) the average sentence score, or (3) the average text score. It may seem logical to look at the average text scores as we are interested in the level of a text. This may, however, not be the best option. There could be a situation in which a text could be given a score that corresponds to that of a Level A text, yet however contains a sentence of Level C. A learner with reading comprehension level A would not be able to understand the whole text. Whether the inclusion of *difficult* sentences in an overall *easy* text is a serious problem might be source for a research amongst learners and teacher. We will get back to this in the discussion. For now we will decide to use the second approach for this reason; if the situation occurs in which a word is too difficult in an overall easy sentence it is easy to give the learner an explanation (or translation into L1) of this difficult word. Again, whether this is the right choice is arguable and should be subject to further research.

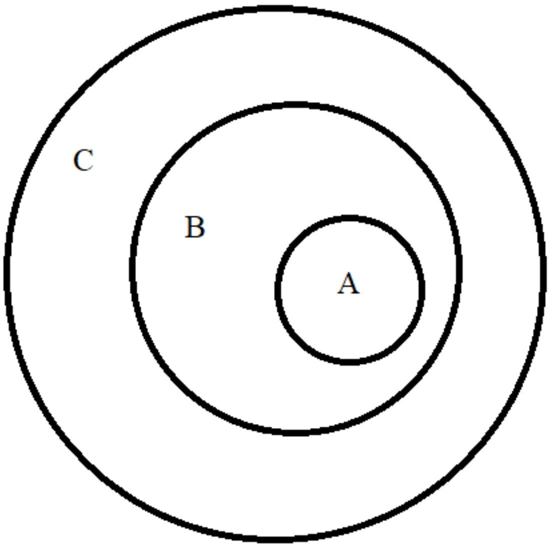

Figure 5 Interpretation of the Levels compared to each other

Figure 5 represents the interpretation of the levels this research will employ. This means that a Level C can contain every possible text, Level B contains both B and A texts, and Level A only contains level A texts. This interpretation is guided by the idea that the reading comprehension of a Level C learner allows this learner to read every possible text. The problem that this interpretation may cause is that a Level A text is wrongfully labelled a Level C text. This, however, will not be a problem for our Classroom-MT because we are only interested in employing simplifications to a Level C text back to a Level B or a Level A text. We are mainly interested in excluding difficult things. It will, however, be beneficial to employ an algorithm that will choose the lowest possible Level that may come out of the calculations. If the logical formulas presented below identify a text as both Level B and Level A, this algorithm will then choose to classify the text as a Level A text.

2.3.1 Level A Text

According to the CEFR Reading Comprehension guidelines, a Level A text should be a short text that uses easy-access language and is recognizable as well as predictable. Easy-access language means easy and high-frequency vocabulary that corresponds to the vocabulary learned in the early stages of L2 acquisition. It also means the text should be short and therefore not contain many words overall. Recognizable and predictable language means shorter sentences. A typical level A text would therefore be short overall, with short sentences and high-frequency vocabulary.

According to the components that determine developmental sequences a level A text should therefore correspond to the following thresholds:

1. The score for *Perceptual Salience* should be HIGH.
    - This means that the score for perceptual salience should be above a certain value. We use the variable xa for the threshold. And PS for perceptual salience. This results in the following formula.
        - PS > xa
2. The score for *Semantic Complexity* should be LOW.
    - This means that the score for semantic complexity should be below a certain value. We use the variable ya for the threshold and SC for semantic complexity. This results in the following formula.
        - SC < ya
3. The score for *Morphophonological Regularity* should be LOW.
    - We use the variable za for the threshold and MR for morphophonological regularity. This results in the following formula.
        - MR < za
4. The score for *Syntactic Category* should be LOW.
    - We use the variable va for the threshold and SCa for syntactic category. This results in the following formula.
        - SCa < va

5. The score for *Frequency* should be HIGH.
    - We use the variable wa for the threshold and F for Frequency. This results in the following formula.
        - $F > wa$

Thus a level A text should conform to the following logical formula for every sentence in the text:

$PS > xa$

$\wedge SC < ya$

$\wedge MR < za$

$\wedge SCa < va$

$\wedge F > wa$

2.3.2 Level B Text

According to the CEFR Reading Comprehension guidelines a Level B text can be longer than a Level A text but should still contain every-day, high-frequency language. It may express certain viewpoints. On the basis of this we can conclude that a Level B text can be longer than a Level A text, and can therefore contain more words overall. Sentences and text should still be shorter than most Level C texts because it should still be every-day and high-frequency language. An intermediate learner will now have a larger vocabulary at level B1 or B2, but is still somewhat limited in the words that can be used in texts. A typical Level B text would therefore be between the length of a Level A and Level C text, with an average amount of words per sentence and use words that are known at this level.

According to the components that determine developmental sequences a Level B text should therefore correspond to the following thresholds compared to a Level A text: It should allow for lower perceptual salience (xb), higher semantic complexity (yb), higher morphological regularity scores (zb), a higher score for syntactic category (vb), and lower frequency (wb).

Thus a level B text should conform to the following logical formula for every sentence in the text:

$$(PC > xb) \wedge (xb < xa)$$
$$\wedge (SC < yb) \wedge (yb > ya)$$
$$\wedge (MR < zb) \wedge (zb > za)$$
$$\wedge (SCa < vb) \wedge (vb > va)$$
$$\wedge (F > wb) \wedge (wb < wa)$$

2.3.3 Level C Text

According to the CEFR Reading Comprehension guidelines a Level C text can be long and complex and may use different styles. It can be specialised and abstract. This is our entry-level text for the machine translator. Long and complex means that there is no restriction on the length of text itself (long) and also no restriction on the length of the sentences and words (complex). This also covers specialised and abstract language, which would mean that non-frequent vocabulary and sentence structures can be used.

According to the components that determine developmental sequences a Level C text should therefore correspond to the following thresholds compared to a Level B text: It should allow for lower perceptual salience (xc), higher semantic complexity (yc), higher morphological regularity scores (zc), a higher score for syntactic category (vc), and lower frequency (wc).

Thus a level C text should conform to the following logical formula for every sentence in the text:

$$(PC > xc) \land (xc < xb \land xb < xa)$$

$$\land (SC < yc) \land (yc > yb \land yb > ya)$$

$$\land (MR < zc) \land (zc > zb \land zb > za)$$

$$\land (SCa < vc) \land (vc > vb \land vb > va)$$

$$\land (F > wc) \land (wc < wb \land wb < wa)$$

3 How to Create a Classroom MT

3.1 Machine Translation Models

This section will introduce two different machine translation models. Firstly, the phrase-based model will be introduced. This model is relatively unaware of the grammatical qualities of the input, it is not based on syntax but largely build upon probability scores that depend on frequencies in the target language. And secondly this section will introduce the tree based model, which is more syntactically aware of the input. It receives as input a parse tree and works from this extra information and (also) probability scores, towards a grammatically correct translated text. Both these models have many different sub-models that are built upon the notions presented in both phrase-based and tree-based models. The sections following these introductions will dive into the uses of these particular models in monolingual machine translation especially.

Koehn (2018) defines phrase-based models as machine translation models that see phrases as atomic units. He adds that they are opposed to, for instance, word-based models that see words as their atomic units. Koehn proposes there are two main advantages to phrase-based models (compared to word-based models). The first advantage is that local context can be used to aid the translation process. And secondly, they can also handle non-compositional phrases.

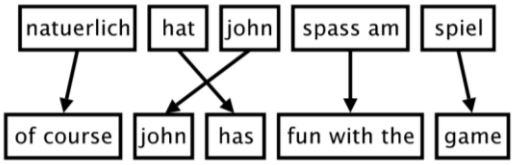

Figure 6 Example of Phrase Based Model (Koehn, 2018)

An example of a machine translation using phrase-based modelling can be seen in Figure 6 where a German sentence is translated into an English sentence. According to Koehn there are three stages in a phrase-based model; (1) the sentence in the foreign language (here German) is segmented into phrases, then (2) these phrases are translated into another language (here English), and lastly (3) the translated phrases are reordered to make a grammatically correct translated sentence (here again English).

Most phrase-based models use statistical models to choose the translation of the phrases. Koehn gives the following example presented in Table 7. Translations for the German word *natuerlich* are given a probability score and the English phrase *of course* is then chosen because it has the highest probability. These scores are based on the word, but also on where the word that has to be translated occurs in this particular moment. It is possible that the translation *naturally* may get the highest score in a different situation.

Table 7 Examples of probability scores for the German phrase "natuerlich" (Koehn, 2018)

Translation	**Probability**
of course	0.5
naturally	0.3
of course ,	0.15
, of course ,	0.05

Koehn adds that phrase-based models are not limited to constituents like NP, VP and others. They can also translate "*non-constituents*". An example of a pair that is not a constituent is found in Figure 6 where the German phrase *spass am* translates to the English phrase *fun with the*. Koehn continues by saying that experiments even show that limitation to these constituents hurts the quality of the translations. He explains that all of the translations done by phrase-based models are based on probability scores created by teaching the model on parallel corpuses and the scoring is based on relative frequency in these corpuses.

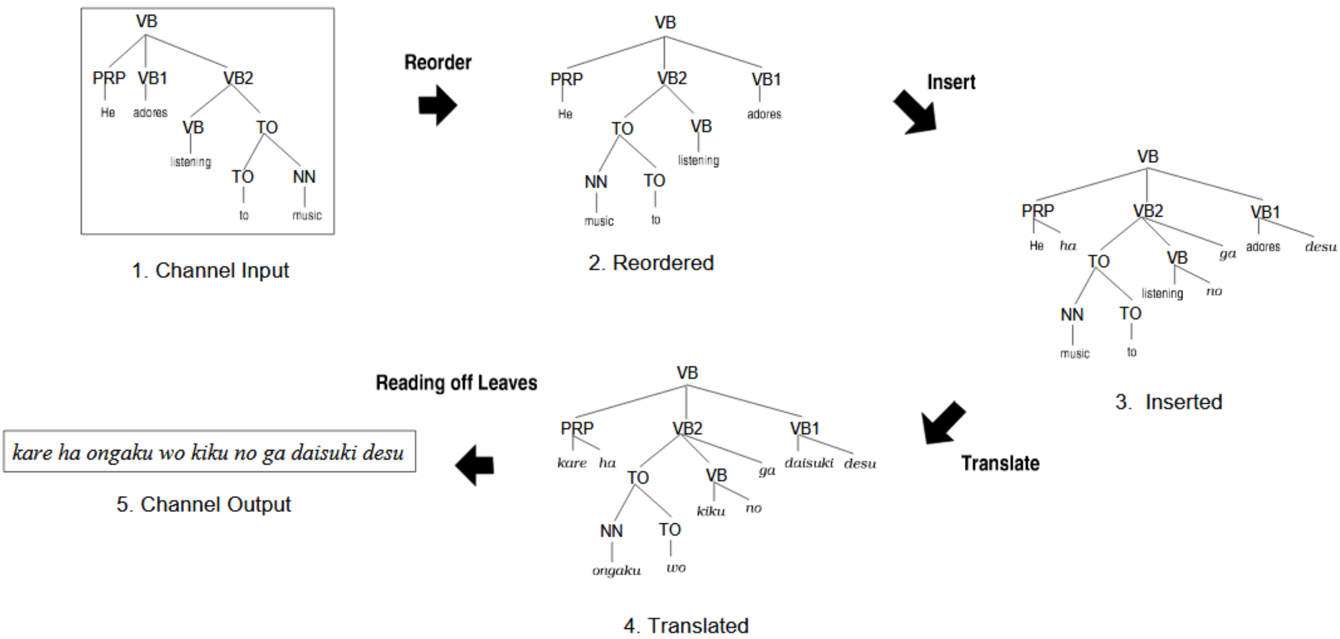

Figure 7 Example of Tree-based Translation Model presented by Yamada & Knight (2001)

The second machine translation model we will discuss, is the tree-based model. A typical example of this is created by Yamada and Knight (2001). They specify that although phrase-based models use phrase classes and their positions in the sentence, they fail to model "structural or syntactic aspects of the language" (p.1). An example of a tree-based model can be found in Figure 7. Here the sentence *He adores listening to music* is translated into the Japanese equivalent *kare ha ongaky wo kiku no ga daisuki desu*. The biggest difference between the phrase-based and tree-based models is how the input is presented and then handled (Yamada & Knight). The tree-based model accepts a parse tree as input and then performs the following operations on them (see Figure 7):

- Reorder the child nodes (1 to 2)
  - The following three reordering operations are done; PRP-VB1-VB2 is changed to PRP-VB2-VB1, TO-NN is changed to NN-TO, and VB-TO is changed to TO-VB.
  - Yamada and Knight only change the sequence of child nodes because they believe they are the ones that influence reordering. This seems to work well because the chance of a sequence remaining grammatical is bigger if a node is not taken from its parent node. They further explain that the amount of possible reorderings within a node is based on the amount of children the node has; “a node with N children has N! possible reorderings” (p. 2). So how to decide which reordering to choose? Each reordering is given a probability score based on a probability formula that itself bases its results on a training corpus (see Yamada & Knight for the formula, p. 3-4).
- Insert extra words at each node (2 to 3)
  - The following *insert* operations are done; *ha* is added to the end of PRP, *ga* is added to the end of VB2, *no* is added to the end of *small* VB, *ga* is added to the end of VB2, *desu* is added to the end of VB1.
  - Yamada and Knight describe that an insertion can be made on the right-side of a node, the left-side or nowhere. This is decided for every node. They continues by saying that, what to insert or where to insert it, is all based on probabilities. These probabilities are calculated based on a training corpus with more than 2000 English parse trees and their

equivalent Japanese sentences. (Formulas can be found in Yamada and Knight, 2004)

- Translate the leaf words (3 to 4)
    - This translation is only based on the word because all the syntactical work has already be done in the *reorder* and the *insert* step. Yamada and Knight state that "no context is consulted" in the translation of the leafs (p. 3).

The output that is generated is in the form of a string, so the parse tree is only needed inside the model. Yamada and Knight state that the biggest advantage of the *reorder* operation is that it accounts for languages with different word orders (SOV, SVO, etc.).

From these two examples the following can be concluded. Both models use statistical ways to choose what to change in the input sentence, and where to change it. Both models use training corpuses to find the probabilities (these are mainly two-sided corpuses with both languages present). It seems the difference between the two translation models presented above is most visible in the following two things. Firstly, the order in which the operations are done. The phrase-based model firstly divides its input in segments, then translates, and then reorders the sentence, while the tree-based model starts with the reordering, then inserts, and ends with the translation. Then secondly, the input to the models is different and therefore the information known by each model is different too. What is done with the information given to the model changes the way in which they translate. Tree-based models have more syntactical information then phrase-based models and therefore seem to know more in terms of sentence structure. They use the way in which the nodes are connected to each other (parents and children) to make sure that grammaticality is upheld. This seems to be an advantage. Phrase-based models, however, seem to be more flexible in translation, as they are not restricted to constituents which has been proven to be beneficial in translation.

The following section will dive into the use of these types of translation models in monolingual machine translation. We will try to answer whether one of the two models works better in the simplification of language or if both models serve their own purpose in monolingual machine translation.

3.2 Simplification Strategies

For the creation of a Classroom-MT we are interested to see how machine translation works within the realms of one language. Monolingual machine translation uses sentence simplification in its translation process. Sentence simplification is used to make people read texts more easily by shortening long sentences and making complex sentences simpler (Zhu, Bernhard & Gurevych, 2010). These simplifications are then used in different environments such as for people with disabilities, or, as is the focus of this paper, for non-native speakers (Zhu et al.). According to Zhu and colleagues there are two main reasons for a sentence to be classified as difficult; syntactical complexity or lexical complexity. This calls for two different forms of sentence simplification; syntactical simplification and lexical simplification (Carroll et al., 1999).

Table 8 Simplification operations

1.	Splitting
2.	Dropping
3.	Reordering
4.	Substitution

Zhu and colleagues identify the above set of common simplification operations (Table 8). Firstly, (1) splitting is the division of a long sentence into shorter sentences. Secondly, (2) dropping is the act of removing ‘unimportant’ parts to make a sentence or text more to-the-point. This can be done in two ways, dropping complete sentences or dropping parts of sentences. Thirdly, (3) reordering can also be done in two ways. It can change the order of the splitted sentences (Siddharthan, 2006) or it can change the order of different parts within a sentence (Watanabe et al., 2009). These are all three syntactical operations. Lastly, (4) substitution is the replacement of difficult words or phrases with a more common or simple synonym. (Zhu et al.). This last one is a lexical simplification.

How does this tie in to the CEFR levels we have seen in the first part of this paper. The focus of the CEFR is the language learner. They make use of a “can-do” attitude (Baldwin & Apelgren, 2018) where they are focused on the accomplishments of the language learner ,and what they are capable of doing in any part of the language learner process. Zhu and colleagues are focused on the mechanics of a text. In a certain way they ask themselves the question what the “can-do’s” of a text are. The idea behind this to how a text can be manipulated to aid a certain group of people, here mostly language learners. The use of sentence simplification, in the support of language learners, can, therefore, be described as a way to match the can-do’s of a text to the can-do’s of a language learner. If they match the language learner will be able to read and comprehend the text. As mentioned by Carroll and colleagues there are ways to do this, lexically and syntactically. We have established that learners go though a series of developmental sequences in both lexical and syntactical components of the L2. They will start with short texts with easy vocabulary, and move on to longer texts with many difficult words. They also go through stages of morpheme learning and question forming. So we will need both syntactical and lexical simplification methods to be able to find the right simplification of text.

The following sections will introduce some of the state-of-the-art simplification techniques in both syntactical simplification (section 3.2.1) and lexical simplification (section 3.2.2). It will give the reader an insight into the different techniques used in monolingual machine translation.

3.2.1 Syntactical simplification

As mentioned above there are three main syntactical simplification operations; splitting (3.2.1.1), dropping (3.2.1.2) and reordering (3.2.1.3). They will be presented separately.

3.2.1.1 Splitting

This section will start with a corpus analysis done by Petersen and Ostendorf (2007). This is done to give the reader an indication about the important characteristics of split sentences (section *Corpus Analysis*). After that, this section will give an example of splitting given by Zhu and colleagues to see how splitting works (section *Example*). And after that a splitting model described by Siddharthan will be presented to show what the underlying techniques are for splitting (section *Splitting Model*).

Corpus Analysis

Petersen and Ostendorf (2007) did a corpus analysis of a two-sided corpus with original sentences and their manually simplified counterparts. The analysis into split and unsplit sentences shows that sentences that will be split are longer (22.3 words) on average than sentences that will not be split (15.0 words). They add that, because splitting will reduce syntactical complexity, it is clear that sentences that will be split have certain syntactic features that unsplit sentences do not have. Petersen and Ostendorf used a C4.5 system to find these features (2007). C4.5 is decision tree learner that can build a classifier for a feature you are interested in. Petersen and Ostendorf used C4.5 to build a classifier for split and unsplit sentences. They looked at three things; length of the sentence in words, the number of different parts of speech (POS) *(adjectives, adverbs, coordinating conjunction, IN (subordinating conjunction), determiners, (proper) nouns, pronouns, and verbs),* and number and average length of several syntactical components of a sentence *(S, SBAR, NP, PP, VP).*

Two of the abbreviations in the above section may be unknown and will be explained here; IN and SBAR. IN is a subordinating conjunction (Eisner, 2018). SBAR is a clause that starts with such a subordinating conjunction (Eisner) and is used as a causal complement (Rooth, 2004) (A causal complement is a complement to a verb (Rooth)). In conclusion, SBAR is a clause that is introduced with an IN that stands in relation to a preceding verb. This IN can also be empty as can be seen in Figure 8.

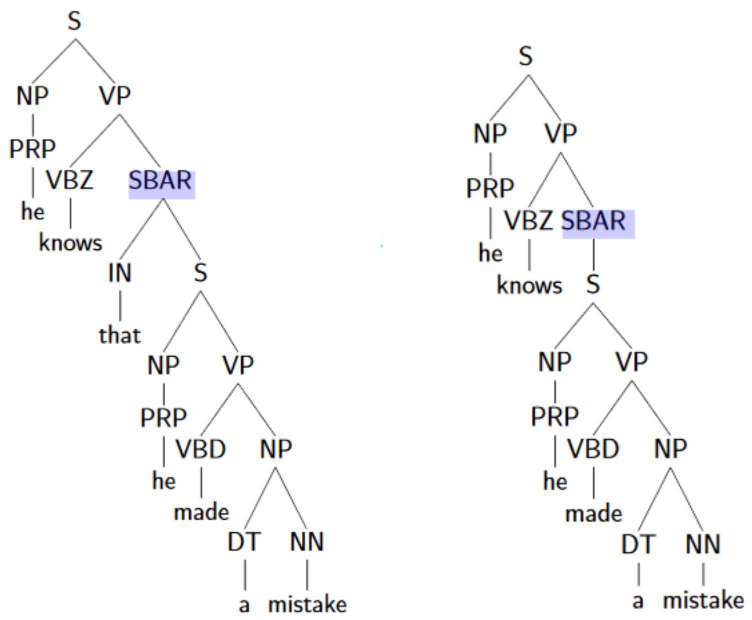

Figure 8 Two sentences with SBAR (Rooth, 2004) (highlight added)

Results of Petersen and Ostendorf (2007), showed that most split sentences had an original length of more than 24 words. Most unsplit sentences were shorter than 19 words. Some sentences that were indeed longer than 24 words but had a small average NP length (<= 1.4 words) were not split. Also, the number of nouns, pronouns, verbs, determiners and verb phrases made a significant impact on the choice of splitting. The number of S and SBAR were not common features that accounted for splitting.

Example

Zhu and colleagues (2010) give the following example for splitting. The example is given here to give the reader an idea of how splitting, and the splitting models presented further on in this section, can be used to simplify a sentence. As we just read most split sentences had an original length of 24 words (Petersen and Ostendorf), and therefore this sentence may seem too short to be split. This is, however, given as an example.

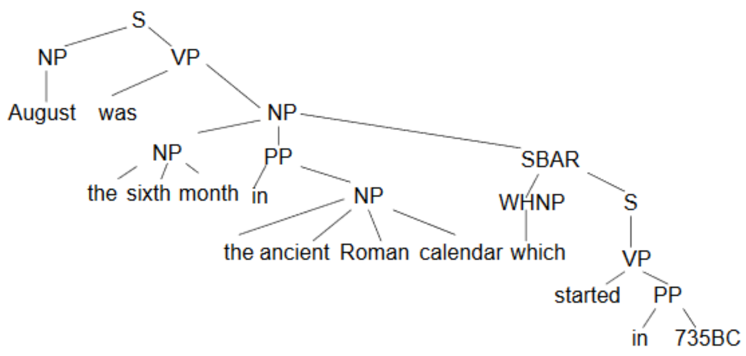

Figure 9 Original sentence (Zhu et al, 2010)

Figure 9 gives the original sentence. This sentence seems quite complicated for a beginner learner. The task is to split the sentence and create two shorter sentences. Zhu and colleagues further simplify the simplification operation by dividing the task into two different subtasks. The first subtask is called *segmentation* and is followed by the second subtask *completion* which is presented later. Segmentation decides whether a sentence should be split at all and where that would then have to happen. Whether a sentence should be split at all is based on the length of the sentence (Zhu et al.) Then an algorithm used by Zhu and colleagues looks for a split boundary word, which is the word *which* in this case. The search for a split boundary word is done in the knowledge that there are many different splitting points possible in a sentence (Zhu et al.). Zhu and colleagues explain that the choice for a split boundary word is based on which one has the highest probability. The probability score is based on many different things. Firstly Zhu and colleagues devised an algorithm that calculates the constituent of word. So for example *which* seems to be WHNP, but the algorithm gives back SBAR because that is where the split will be made. Another feature that is used in the calculation of probabilities is something calles iLength. Zhu and colleagues explain that this is the length of the complex sentence compared to the average simplified sentences in the training data set. So the features that are of importance to Zhu and colleagues are (1) the word in question, (2) what its constituent is and (3) the length of both the complex sentence and simplified sentences in general. Figure 10 shows the possible result of the segmentation task presented by the researchers.

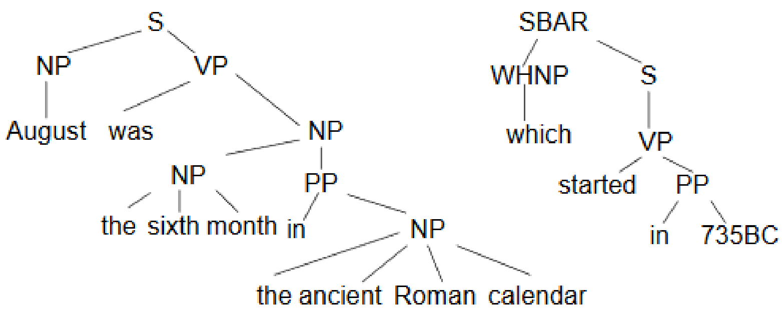

Figure 10 After segmentation (Zhu et al, 2010)

After segmentation they proceed with the second subtask; *completion*. This task is used to make the split sentences, given after segmentation, grammatical and thus complete. The algorithms used by Zhu and colleagues determine whether the first word should be dropped and if that happens what words should be added. Why this is important can be seen in Figure 11. The word *which* (which headed the second sentence in Figure 10) is dropped and replaced by the dependent NP from sentence 1; *the ancient Roman calendar*. Now both sentences are grammatically correct and complete.

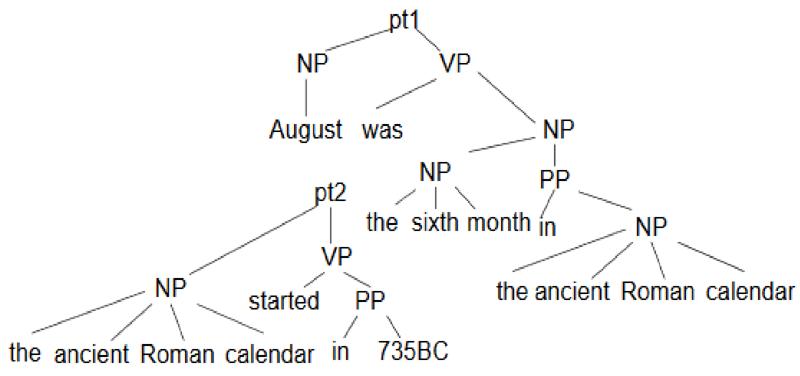

Figure 11 After completion (Zhu et al, 2010)

This paper will return to these sentences because they can be further simplified by using dropping techniques that are discussed further on in this paper. What is interesting to note here is that the sentence presented above is split on SBAR which was deemed as a non-common feature by Petersen and Ostendorf (2007). It is important to understand that Petersen and Ostendorf were looking at the characteristics of sentences that were split and the number of SBARs were not a common characteristic of split sentence. This, however, does not tell us if an SBAR can be a good syntactic constituent to split a sentence on. Zhu and colleagues, are not looking at the amount of SBARs but whether this particular SBAR was a good place to split the sentence. That does not take away from the fact that it is surprising that the number of SBARs is a non-common characteristic. An SBAR indicates an embedded sentence, and a high amount of SBARS seemed to indicate that a sentence is more embedded. Maybe the number of SBARs does not indicate embeddedness, or maybe the embeddedness of a sentence does not signify complexity.

Up until now we have seen the most common characteristics of split sentences that were presented by Petersen and Ostendorf in number and averages. Additionally, we have seen an example of how a sentence is split. This example demonstrated that even though the corpus analysis showed that the number of SBARs is not an indication of whether a sentences will be split, it can be the best place for splitting. The following section will illustrate how sentences are split on the basis of rules and constraints, and what splitting implies for grammaticality.

Splitting Model

Siddharthan (2006) describes syntactic simplification as the act of lowering a text´s grammatical complexity to increase readability. An important thing to be aware of, according to Siddharthan and many other researchers in this area, is that while reducing the grammatical complexity it is vital to not lose sight of the information that is conveyed in the original text. The challenge lies in the fact that while replacing some syntactic parts of a sentence the content and meaning remains the same but the result is still grammatical.

According to Siddharthan there are three stages in syntactical simplification; analysis, transformation and regeneration (Figure 12).

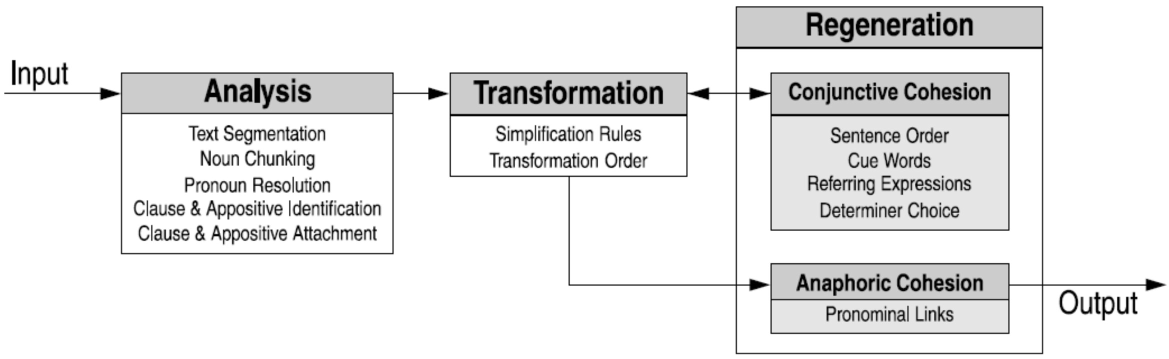

Figure 12 Siddharthans "architecture for syntactic simplification" (Siddharthan, 2006)

In the analysis stage the input text is analysed and tagged for future reference. By doing this the system used in the transformation stage “knows” the syntax of the input text. Siddharthan uses many different known resources and toolkits to do this (see Siddharthan, 2006, p. 80 for more information on these resources and toolkits).

The transformation stage uses rules that were handmade by Siddharthan for these purposes. An example of these kinds of rules can be found in Figure 13 below. The gist of this rule is that if a relative clause called RELPR Y, is attached to a noun phrase W, it is possible to create a new sentence with W and Y.

$$\langle \mathtt{s} \rangle \mathtt{VW}^{n}_{NP}\mathtt{X}[_{RC}\mathtt{RELPR}^{\#n}\mathtt{Y}]\mathtt{Z.}\langle /\mathtt{s} \rangle \longrightarrow \begin{array}{ll} (i) & \langle \mathtt{s} \rangle \mathtt{V\ W\ X\ Z.}\langle /\mathtt{s} \rangle \\ (ii) & \langle \mathtt{s} \rangle \mathtt{W\ Y.}\langle /\mathtt{s} \rangle \end{array}$$

Figure 13 Rule for syntactic simplification (Siddharthan, 2006)

Suppose you have an original sentence like sentence (a). You can create two different simplified sentences by using the rule in Figure 13 (b and c).

(a) An Artificial Intelligence student, *who lives in Utrecht*, passed all his first-year courses.

(b) An Artificial Intelligence student passed all his first-year courses.

(c) An Artificial Intelligence student lives in Utrecht.

Siddharthan adds another stage because the issue with only using the above stages is that there might occur problems with sentence order and meaning change. Sentence (d) can be simplified into the two following sentences (d’ & d’’), and then further simplified into the three sentences below that (d.1-d.3). In the original sentence *but she knows it might be important for her future* is linked to *The bachelor student hates making notes*. But if the sentences of the simplification are put in this order the final sentence is (wrongly) linked to *The bachelor student ran a committee in the student organisation*. This problem is a problem of conjunctive cohesion. Another problem also occurs namely a problem of anaphoric cohesion. The pronoun it (in d.3) now refers to *running a committee* and not to *making notes*.

(d) The bachelor student, who ran a committee in the student organisation, hates making notes, but she knows it might be important for her future.

(d’) The bachelor student, who ran a committee in the student organisation, hates making notes.

(d’’) But she knows it might be important for her future.

(d.1) The bachelor student hates making notes.

(d.2) The bachelor student ran a committee in the student organisation.

(d.3) But she knows it might be important for her future.

Siddharthan determines that there are two problems that occur after simplification. He also put two parts in the regeneration stage (see Figure 12). The first part deals with conjunctive cohesion and is called upon repeatedly during the transformation stage. The second part deals with anaphoric cohesion which is only called upon at the end of the transformation stage.

The transformation stage is called upon recursively. Siddharthan explains that it works in the following manner:

1. The transformation stage splits a sentence in two
2. Both sentences are sent to the regeneration stage
3. The regeneration stage deals with conjunctive cohesion issues
4. The two regenerated sentences are pushed on the stack to be transformed (in the order given by the regeneration stage)

Only after the transformation stack is empty the simplified texts is sent to the regeneration stage again. The regeneration stage then handles problems with anaphoric cohesion before outputting the whole text. The regeneration stage posed by Siddharthan uses the constraint satisfaction problem to resolve issues. It would resolve the issues posed in sentence (d) and its simplifications by giving the constraints. Siddharthan explains that the constraints for the sentences above would be that d’ < d’’ makes sure that this is the order in which the sentences are put, and d.1 > d.2 makes sure that these sentences are put in a different order. In this way sentence (d) is simplified to:

> *The bachelor student ran a committee in the student organisation. The bachelor student hates making notes. But she knows it might be important for her future.*

The simplification example by Zhu and colleagues (2010) and the explanation of the technique by Siddharthan (2006) may seem different but they are actually quite alike as to where they split the sentences. The SBAR principle put forward in the Zhu example is very similar to the RELPR principle by Siddharthan. The word *but* (in example-sentence d) is a subordinating conjunction (IN) that is the head of an SBAR (as explained above in the definitions given). The biggest difference between the two approaches is the way they go through their procedures. Zhu and colleagues present two tasks; segmentation and completion. Siddharthan gives three stages; analysis, transformation and regeneration. The segmentation task seems to be relatively similar to the analysis and transformation stages. The biggest difference is probably found between the tasks *completion* and *regeneration*. Both tasks have the goal to make the simplified sentences grammatical, but do it in different ways.

While Zhu and colleagues rely on probabilities and for deciding what word should be dropped (or not) and what phrase (or word) should be added to make a sentence grammatical, Siddharthan uses constraints and rules to make simplified sentences grammatical. He created different constraint for every single simplification rule he made, and even different constraints for different subordinating conjunctions. These constraints are build upon syntactical knowledge of Siddharthan. The question arises whether this is possible language-independently or whether the method by Zhu and colleagues would work better. This is probably linked to how complex a feature is and how many different constraint are needed for the language independent variables presented by Goldschneider and DeKeyser (2001). The advantage is probably that the use of the constraint satisfaction problem is less prone to errors because it has fixed results.

3.2.1.2 Dropping

There are two sorts of dropping that will be discussed in this paper. Firstly, the dropping of complete sentences, and secondly, dropping of parts of sentences. Both of these dropping techniques are used to simplify texts by dropping unnecessary or insignificant parts Petersen & Ostendorf, 2007). This section will discuss both *sentence dropping* and *parts dropping*. Parts dropping is not covered by Petersen and Ostendorf so Zhu and colleagues and the new models introduced here will cover that part.

Sentence dropping

Just like their analysis of the corpus for split sentences Petersen and Ostendorf (2007) determined what characterizes sentences that are dropped. This analysis was only for complete sentence dropping. So the results give an indication about what characteristics a complete sentence has, to be chosen for dropping. The following features were checked by Petersen and Ostendorf.

- Position in the document (sentence number, percentage)
- Paragraph number, first or last sentences in paragraph
- Does this sentence contain a direct quotation?
- Percentage of words that are stop words

- Percentage of content words which have already occurred one/two/three/four/more times in the document

Stop words are usually words that are the most common in a language. And content words are words that signify objects in reality and what they are like or what they do.

Again results were found with the help of a C4.5 system. Petersen and Ostendorf found that direct quotes were often deleted from the text, an especially when the quotation contained more than 70% stop words. They also found that position in the document did seem to matter as well and quotes were often deleted when they were past the twelfth sentence. In general it means that later sentences (past the 35th sentence) were dropped more (Petersen & Ostendorf).

Parts Dropping

<u>Example</u>

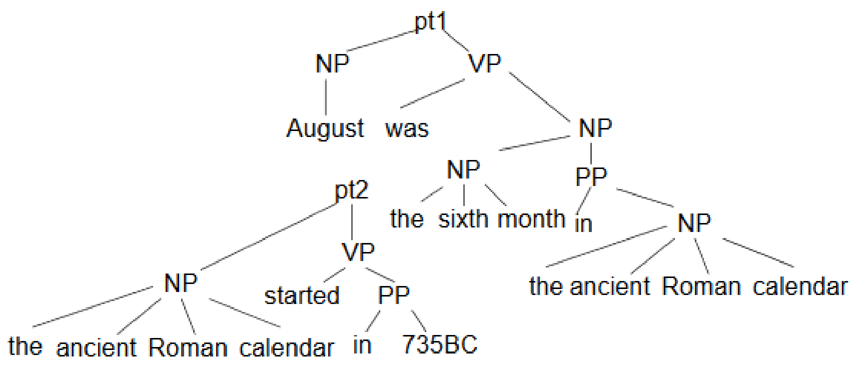

Figure 14 Before dropping (Zhu et al, 2010)

Figure 14 gives the tree structure that we were left with after sentence splitting in the previous section. The dropping algorithm used by Zhu and colleagues gives the following result, namely the dropping of the word *Roman* in both sentence parts. Figure 15 shows what the sentences are like after applying splitting and dropping to then.

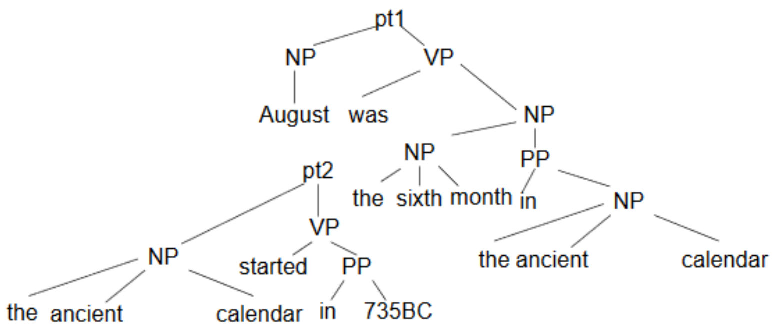

Figure 15 After dropping 'Roman' (Zhu et al, 2010)

As can been seen from the above there is big difference in the approach of Zhu and colleagues compared to Petersen and Ostendorf. Zhu and colleagues drop unnecessary parts of a sentence and Petersen and Ostendorf look into completely dropped sentences. The goal of dropping is to make the text shorter. Petersen and Ostendorf mention that in the past people have tried to use summarization techniques to make a text shorter, these summarization techniques choose a certain number of sentences to create a subset of the total amount of sentences. They however, mention that the problem with these summarization techniques is that, while they may shorten a text, complex sentences could still be in the text. It might be beneficial to do sentence dropping as the first simplification and then move on to different simplification techniques like splitting, dropping parts of sentences, reordering, and substitution. In this way the text that has to be simplified is already shorter and will need less simplification operations. The following section will dive into the second dropping technique; parts dropping.

<u>Parts Dropping Models</u>

This paper will discuss two different kinds of parts dropping models. Both of these have been proven to work quite well. These models were put forward by Knight and Marcu (2002), who have used two types of probabilistic based dropping models. Both these models were used before in other areas of linguistics, so they are not new. They were however, according to Knight and Marcu, never used for compression strategies. Compression is the act of making a sentence smaller than it is. This can be done through splitting, as we discussed before, or dropping certain parts of a sentence.

The first model we will discuss is the *noisy channel model*. Knight and Marcu describe the noisy channel model in the following manner. A noisy channel model comes forth from the idea that if you find a long string you should try to find the hypothetical short string that it once was. You believe someone added additional/optional text to it. This way of working faces the task of finding solutions to three problems which will be discussed in this section. The word *right* is used a few times in these solutions. The *right* answer in these solutions is the answer with the highest probability.

(1) The problem is that you should be able to choose the right (grammatically correct) short string. Knight and Marcu define the first solution as a source model. This source model should be created in such a way that, when given P(s), it gives back the probability that a string *s* was the original short string. When *s* is ungrammatical we would, for instance, want to receive a very low score for P(s).

(2) The second problem is being able to check if a simplification *s* is indeed a version of the longer (complex) sentence *t*. As a solution for this problem Knight and Marcu describe a channel model. This channel model should, when given P(t | s), return the probability that a pair <*s,t*> contains a short string *s* (a simplified sentence) and its expansion long string *t* (the (complex) input sentence). Suppose sentences *s* and *t* (see (i) and (ii)) are the exact same sentence except for the negation '*not*'. Theoretically the sentence is now simplified because a word is dropped. However, the model should return a low score for P(t | s), because a negation is not optional or additional. It is a vital part of a sentence and important information is now lost, and (i) cannot possibly be the right simplification of (ii).

i. Marie is a football player.

ii. Marie is not a football player.

(3) The third problem is finding the right simplification of a long and complex sentence. So the third model that will need to be created according to Knight and Marcu is called a decoder. This decoder will have be to be created in such a way that when a long string *t* is found, it will search for a short string *s* that maximizes P(s | t). Knight and Marcu explain that this is the model/solution that makes use of the two solutions above, because finding the maximum P(s | t) is equivalent to searching for an *s* that maximizes P(s) * P(t | s) (*Bayes Theorem*).

By creating three models you can create different models for different problems you are facing in the production of sentences with dropped parts. Knight and Marcu state that the two issues with dropping are that (1) you want something that is grammatical but also (2) preserves important information. The problem lies in finding the balance between them. An option presented by Knight and Marcu is that you can use the source model to find a grammatical sentence and the channel model to find a sentence that preserves important information. (*For more information on creating a simple noisy channel model see* Knight & Marcu, 2001).

The second model we will discuss is also a probabilistic model called a decision based model. Knight and Marcu describe it as follows. The input model receives a parse tree 't' this is (in our case) a long and complex sentence, and the goal would be to rewrite 't' in such a way that it turns into a smaller tree 's', which is a simplified version of 't'. To do this there is an empty stack and an input list that contains the words that make up the large tree 't'. Each word in the input list receives a label that contains all syntactic constituents in the large tree 't' that start with it. The model goes through a series of steps that all have the aim to reconstruct a smaller tree 's' and all steps work towards that. The decision based model uses 4 operations, presented by Knight and Marcu, to do this, all of which will be explained below.

The first operation is called SHIFT. This operation takes the first word from the input and pushes it onto the stack.

The second operations is called REDUCE. This operation pops a certain number of syntactic trees from the top of the stack. Additionally, the operation combines all these trees into a new tree and pushes the new tree on top of the stack. By doing this the operation derives the syntactic tree of the short sentence 's'.

The third operation is called DROP. DROP deletes words from the input list that conform to the same syntactic constituents. An example: DROP X will delete all words in the input list that X spans in the tree 't'.

The fourth and final operation is called ASSIGNTYPE, which will change the label of the a tree at the top of the stack. It is used to appoint POS tags to words in the eventual compressed sentence. It is important to note that because of this operation it is possible that the POS tags in the compressed sentence are different to the ones in the original sentence. Because of this the decision based model is more flexible than the noisy channel model. It can derive trees whose structure can be very different from the input given to the model.

Figure 16 shows an example by Knight and Marcu of this decision based model. The three trees at the top show the complex tree (t), and the two possible simplifications (s1) and (s2). And the Figure below the trees shows the steps taken to get from (t) to (s2) by using the operation presented above. The nodes K and F are introduced and make it possible for the reordering of the word order. In both (s1) and (s2) the tree under B is dropped. Apparently this part of the sentence was deemed redundant.

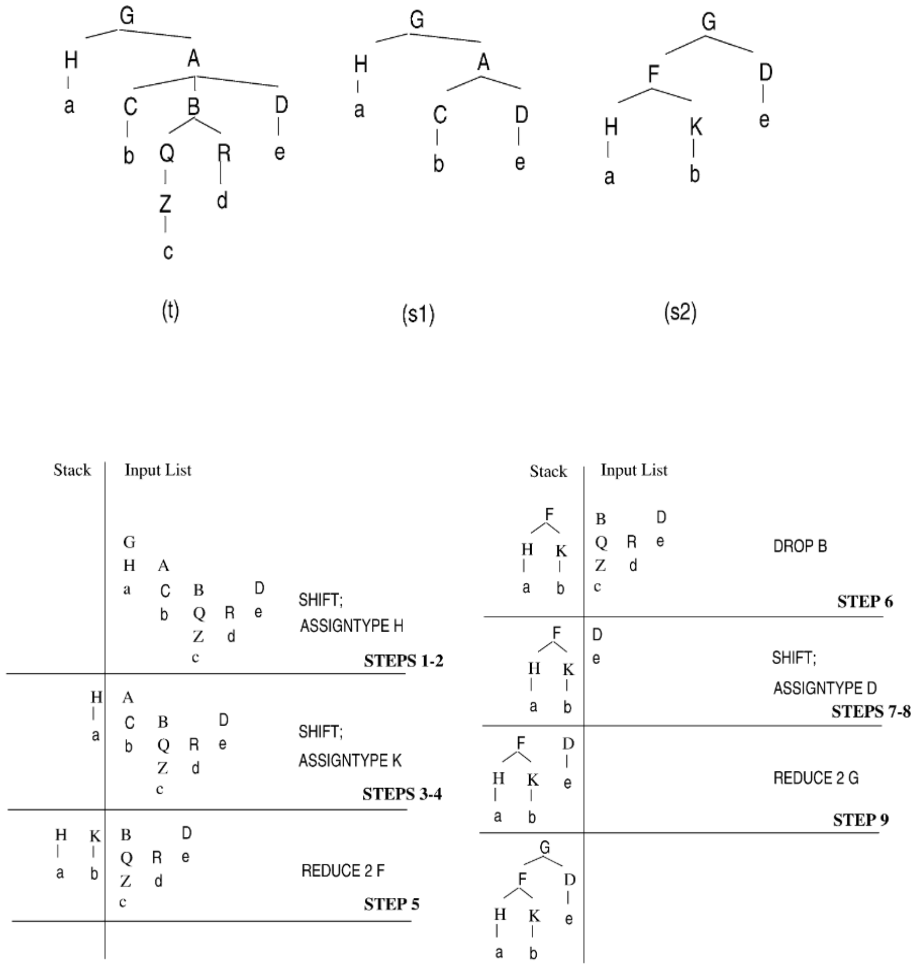

Figure 16 Example of Decision Based Model (Knight & Marcu, 2002)

The advantage of the decision based model compared to the noisy channel model is the ability to go beyond sentence structure because of the ASSIGNTYPE step. It may however, be a more difficult model to implement because of this step especially. It might also make the error rate higher because of the flexibility of the sentence structure.

3.2.1.3 Reordering

Reordering is a major factor in machine translation (Xiong, Liu & Lin, 2006; Wang, Collins & Koehn, 2007; Tillman & Ney, 2003; Zens & Ney, 2003; and many others). Most languages do not have the same word order and therefore machine translators need a good system for the reordering of the translated phrases as they strive to find grammatical translations. As was mentioned in the splitting section of this paper, monolingual machine translation uses the reordering of complete sentences for conjunctive and anaphoric cohesion (Siddharthan, 2006). The regeneration stage provided by Siddharthan is a good example of how the constraint satisfaction problem is used in reordering situations.

This way of working with constraints may work well for reordering because Radford (2004) shows that word order parameters are binary in nature. Radford mentions the Head-Position Parameter (whether a language is head-first or head-last), and the Wh-Parameter (are wh-expression fronted in questions or not). The points is that all these parameters have two choices and every language in the world is either one or the other (Radford). If it would be possible to have list of parameter-settings for all languages and parameters, the task of creating constraints for every language might not be such a difficult task. You then would either switch it on or off for every parameter.

However, Narayan and Gardent (2014) state that (Wubben et al., 2012) have proven that a statistical machine translation model for sentence simplification also works well for reorderings. One would need to feed alignment and probabilities to the model. Narayan and Gardent, therefore, use a phrase-based MT for their lexical substitutions as well as their reorderings. They conclude that this approach results in good “fluency and grammaticality” (p. 437).

3.2.2 Lexical simplification

Just like in the section on syntactical simplification, we would like to introduce an example of what simplified texts will be characterized by. This will be done with the introduction of the *Leesniveau Tool* in the following section. The following section will then give an example of simplified text and what it means. And the finally we will describe the process of simplifying a text by … substitution. Substitution is the fourth, final, and only lexical simplification strategy introduced by Zhu and colleagues (2010).

3.2.2.1 Leesniveau Tool

The *Leesniveau Tool* (Reading Level Test Tool; RLTT) is a free Dutch tool that analyses the language used in webpages and gives a conclusion based on the CEFR levels (Velleman & Van der Geest, 2014). The introduction on the Dutch Wikipedia page on mathematics (wiskunde[4]) is, for instance, classified as C1. This classification gives the makers of webpages a basic indication of their use of language (Velleman & Van der Geest). The RLTT was created to make sure that important webpages (for instance governmental) are understood by the people using these webpages (Velleman & Van der Geest). The average Dutch text reading comprehension level for Dutch native speakers on the CEFR scale is B1. Velleman and Van der Geest explain that this means that they lack the reading comprehension skills needed to understand their own native governmental webpages, because they require at least C1 reading comprehension level. The purpose, according to Velleman and Van der Geest, of the tool is not so much to give an exact classification but to raise awareness. However, Jansen and Boersma compared the RLTT to two commercial test-tools and concluded that the RLTT, although free, was the most effective (Jansen & Boersma, 2013).

[4] https://nl.wikipedia.org/wiki/Wiskunde

Different parameters were used in the RLTT as shown in Table 9.

Table 9 Parameters used in the RLTT

1.	Average number of simple words
2.	Average number of words per sentence
3.	Average number of pronouns in a sentence
4.	Average number of syllables per word
5.	Average number of prepositions per sentence
6.	Number of names and terms

The RLTT uses a database to check the number of simple words in a sentence (1), this database is composed of words known to be simple by users and also words proposed by specialists to be used often in re-translating texts (simplifying) (Velleman & Van der Geest). The average is then taken by dividing the amount of these simple words by the amount of total words in the sentence.

The amount of syllables per word (4) is calculated by looking for separate combinations of vowels in the words (from three vowels together (e.g. aai), to two vowels (e.g. aa), to one vowel (e.g. a)) (Velleman & Van der Geest).

To calculate the average number of prepositions in a sentence (5) the total number of prepositions is divided by the total amount of sentences in the text. To find prepositions the text is searched by comparing the words to a list of prepositions (91 Dutch prepositions) (Velleman & Van der Geest).

All these parameters work together to find the right reading comprehension level of a text. Although only looking for lexical characteristics of a text, it is shown to be highly effective compared to other test-tools. This shows that lexical information is also very important to assess the level of a text, and also to the level of a text in general. And it also shows the characteristics of lexically simplified language. This would make it easier to check if a simplified text adheres to these characteristics.

3.2.2.2 Basic English

A form of lexical simplification is the use of Basic English. This form of English was introduced by Charles Kay Ogden in the 1930s. Basis English is a form of English that is very limited in its number of words (850), giving it a small vocabulary (Wubben et al., 2012). According to Ogden "90 percent of all dictionary entries can be paraphrased using these 850 words" (as cited in Wubben et al.). This means that even with fewer words it is possible to convey the same meaning. This is important in the simplification procedures, because you want to make sure that a less-skilled reader is still being given the same information. Or at least given the same gist of the story. Basic English nowadays is mainly used in the Simple Wikipedia pages, where sentences are less complicated and shorter than in usual Wikipedia pages (Wubben et al.). Zhu and colleagues (2010) provide the following statistics presented in Table 10.

Table 10 Statistics (average sentence length and average word length) of Simple Wikipedia compared to Wikipedia (Zhu et al., 2010)

	Sentence Length	**Token Length**
Simple Wikipedia	20.87	4.89
Wikipedia	25.01	5.06

The sentences on Simple Wikipedia pages are 17% shorter than the sentences on regular Wikipedia pages and the length of the average word is almost the same (Wubben et al.). This shows that a lexical simplification of English is not only concerned with word choice (limited vocabulary) but also with sentence length.

3.2.2.3 Substitution

The lexical simplification posed by Zhu and colleagues will be presented in this section. The goal of substitution is to exchange difficult words for their more simpler synonyms. This can be done using multiple different sources and toolkits. Inui and colleagues

(2003) pose the use of WordNet and dictionaries. Wubben and colleagues (2012) present a word-substitution baseline model. For every noun, verb and adjective in the sentence the word and POS-tag (*computed with the Memory-Based Tagger* (Daelemans et al., 1996, as cited in Wubben et al.)) are taken an fed to WordNet that produces all synonyms of the word. The model then uses the SRILM language model presented by Stolcke (2002, as cited in Wubben et al.) that scores each replacement by probability. This probability scoring is based on the Simple Wikipedia dataset that was mentioned above. The synonym with the highest probability is then put in the place of the difficult word. There is always a chance that a synonym is not found, in that case the difficult word does not change (Wubben et al.).

From the above it becomes clear that for substitution purposes it is easy to find toolkits and datasets. The important thing to understand is that one needs three things to be able to substitute a word; a (1) POS-tagger, (2) a dataset with synonyms, and (3) a dataset to check for probabilities in simplified language.

3.3 So what now?

We have seen syntactical and lexical simplification methods that have been proven to work well in (monolingual) machine translators. But the aim of this paper was not necessarily to give an overview of the methods that are available for simplifying language. The goal of this paper has been to give a starting point for the creation of a Classroom-MT. In the first part of this paper we have established that there are certain components of language that determine when a certain language phenomenon is learned. We have also established that the length of a text and the choice of words influence the *Level* we attribute a text to. The interesting part would now be how we can combine these things we established in the first part (chapter 2) of this paper and the simplification methods we have seen in part two (chapter 3). This will be further developed in the conclusion of this paper.

We have seen that many simplification methods use training data to be able to calculate probabilities for certain language components. Most choices within simplifications are based on probabilities. Siddharthan (2006) probably being one of the only exciting ventures into a different way of simplification choices; the use of the constraint satisfaction problem in the simplification process. Some models use phrase-based machine translation and some use tree-based machine translation. All different simplification strategies have different advantages and disadvantages for either of the two models. In the creation of a model that develops a Classroom-MT both forms will probably have their place.

4 Conclusion, Discussion & Further Research

The aim of this study was to answer the following questions:

Research Question: How can we create an MMT that, when given the reading comprehension level of the language learner and a certain text, produces a simplification of the given text that is in accordance with the given level?

a. How can we classify L2 learners in terms of reading comprehension?
b. What simplification methods and language models are used in monolingual machine translation?
c. How can we connect learner levels to simplifications in monolingual machine translation?

Up until now we have answered the sub-questions (a) and (b) in section 2 and 3 of this paper. We have found that learners can be classified using the CEFR scales and the developmental variables posed by Goldschneider and DeKeyser (2001). We have also found that these variables can easily be formalised to create different text levels. What the values will have to be remains open for further research. We suggest the use of a C4.5 model (used by Petersen and Ostendorf, 2007) in finding the probabilities of these variables in a training data set with for all different levels. A database would have to be created per level.

We have also seen that machine translation models are mainly phrase-based and tree-based and are also used in monolingual machine translation models. We have seen that syntactical simplification uses both of these commonly. We have seen splitting models that make use of trees, phrases and clauses. We have seen dropping models that use phrases as well as trees. And reordering models that rely on phrase based statistical machine translation approaches. Lexical simplifications have been shown to be more probabilistic based.

The subquestion (3) still needs to be answered. We now know a way in which language development can be formalised (subquestion 1) and we know the common ways in which language can be simplified (subquestion 2).

At this moment there are no simplification models that deal with the first four variables posed by Goldschneider and DeKeyser directly.[5] One could use the simplification methods to simplify a text and then calculate the value of the variables and see what the level of the text is. The variables will then be used as a checking method. However, it might be possible to create simplification methods that deal especially with (1) perceptual salience, (2) semantic complexity, (3) morphophonological regularity, and (4) syntactic category. These simplifications would need to be able to differentiate between Level A and Level B simplifications.

Further research should focus on finding the values of the variables used in these logical formulas and, check if these variables adhere to the language levels. This paper proposes the following method.

[5] The fifth variable, Frequency, is covered in lexical substitution.

To check if the texts that are created are the *right* kind, there are two things that need to be checked:

i. Whether the text is (1) grammatical and (2) whether it contains the information conveyed in the original text.

ii. Whether the level of the text is the right level.

   a. Possibly: Whether the text is in the "reach" of a Level A, B, or C text according to our logical formulas.

We propose the use of two different groups for these two parts. For the first check (i) we propose to use foreign language teachers to check for grammaticality and information. For the second check we propose to use actual language learners (opposed to language teachers who were used in previous research) because they will be able to (1) communicate whether they understood a text and (2) asking the right questions will also give an indication of whether they did actually understand the text. It will also be important to investigate whether it will matter if we choose sentence score over word score as mentioned in section 2.3.

Another step that can be taken in further research is to check how the models proposed in this paper will compare to more recent forms of machine translation, especially *Neural Machine Translation*, in the creation of a Classroom-MT. A thorough introduction of Neural Machine Translation is beyond the scope of this paper, however, this form of machine translation has received a lot of attention lately and should at least be mentioned in this paper. Although Neural Machine Translation works better than the MTs presented in this paper, Wang and colleagues (2016) very clearly explain the problem with this form of machine translation in text simplification. They state that a single Neural Machine Translator is "not able to handle different text simplification operations" (p. 4271). And since we have established that many operation are needed for text simplification work will have to be done before Neural Machine Translation will find its place in text simplification.

The question that remains is whether it is possible to create such a Classroom-MT. The answer given by this paper is that it seems theoretically possible to do so, both from a L2 language learner point of view, as well as the monolingual machine translation perspective. Whether it will work in the way expected, and will actually help language teachers and learners to advance in their language levels, remains to be seen.

5 References

Baldwin, R., & Apelgren, B. M. (2018). Can Do and Cannot Do–CEFR inspired examination and assessment in a Swedish higher education context. *Apples: Journal of Applied Language Studies*, *12*(2).

Carroll, J., Minnen, G., Pearce, D., Canning, Y., Devlin, S., & Tait, J. (1999). Simplifying text for language-impaired readers. In *Ninth Conference of the European Chapter of the Association for Computational Linguistics*.

Collins, L. (2002). The roles of L1 influence and lexical aspect in the acquisition of temporal morphology. *Language learning*, *52*(1), 43-94.

Council of Europe. (2001). Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). Retrieved from https://rm.coe.int/1680459f97

Council of Europe. (2018a). *The CEFR Levels*. Retrieved from https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions

Council of Europe. (2018b). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR).* Retrieved from https://www.coe.int/en/web/common-european-framework-reference-languages

Council of Europe. (2018c). *Global scale - Table 1 (CEFR 3.3): Common Reference levels*. Retrieved from https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale

Council of Europe (2018d). Self-assessment grid - Table 2 (CEFR 3.3) : Common Reference levels. Retrieved from https://www.coe.int/en/web/common-european-framework-reference-languages/table-2-cefr-3.3-common-reference-levels-self-assessment-grid

Eisner, J. (2018). *Nonterminals and Preterminals in the Penn Treebank II* [PDF file]. John Hopkins University, Course in Natural Language Processing. Retrieved from http://cs.jhu.edu/~jason/465/hw-parse/treebank-notation.pdf

Gass, S. (1982). From Theory to Practice. in M. Hines and W Rutherford (eds.): *On TESOL* '81. Washington, DC: TESOL.

Garcia, I., & Pena, M. I. (2011). Machine translation-assisted language learning: writing for beginners. *Computer Assisted Language Learning*, *24*(5), 471-487.

Gershgorn, D. (2016). *Google says its new AI-powered translation tool scores nearly identically to human translators*. Retrieved from https://qz.com/792621/googles-new-ai-powered-translation-tool-is-nearly-as-good-as-a-human-translator/

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the “natural order of L2 morpheme acquisition” in English: a meta-analysis of multiple determinants. *Language learning*, *51*(1), 1-50.

Grabe, W. (2014). Key issues in L2 reading development. In *CELC Symposium Bridging Research and Pedagogy* (pp. 8-18).

Harsch, Claudia & André Alexander Rupp (2011) Designing and Scaling LevelSpecific Writing Tasks in Alignment With the CEFR: A Test Centered Approach, Language Assessment Quarterly, 8:1, 1-33

Inui, K., Fujita, A., Takahashi, T., Iida, R., & Iwakura, T. (2003, July). Text simplification for reading assistance: a project note. In *Proceedings of the second international workshop on Paraphrasing-Volume 16* (pp. 9-16). Association for Computational Linguistics.

Jansen, C., & Boersma, N. (2013). Meten is weten? Over de waarde van de leesbaarheidsvoorspellingen van drie geautomatiseerde Nederlandse meetinstrumenten [*About the worth of readability-predictions of three automated Dutch measuring instruments*]. *Tijdschrift voor Taalbeheersing* [Magazine for Mastery of Language], *35*(1), 47-62.

Knight, K., & Marcu, D. (2002). Summarization beyond sentence extraction: A probabilistic approach to sentence compression. *Artificial Intelligence*, *139*(1), 91-107.

Koehn, P. (2018). Machine Translation: Phrase-Based Models [PowerPoint slides]. Retrieved from http://mt-class.org/jhu/slides/lecture-phrase-based-models.pdf

Laver, J., & John, L. (1994). *Principles of phonetics*. Cambridge University Press.

Lightbown, P. M., & Spada, N. (2013). *How languages are learned 4th edition-Oxford Handbooks for Language Teachers*. Oxford university press.

McNamara, T. (2011). Managing learning: Authority and language assessment. *Language Teaching*, *44*(4), 500-515.

Meisel, J. (1987). Reference to past events and actions in the development of natural second language acquisition. *First and second language acquisition processes*, 206-224.

Narayan, S., & Gardent, C. (2014). Hybrid simplification using deep semantics and machine translation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Vol. 1, pp. 435-445).

Niño, A. (2009). Machine translation in foreign language learning: language learners' and tutors' perceptions of its advantages and disadvantages. *ReCALL*, *21*(2), 241-258.

Petersen, S. E., & Ostendorf, M. (2007). Text simplification for language learners: a corpus analysis. In *Workshop on Speech and Language Technology in Education*.

Pienemann, M., Johnston, M., & Brindley, G. (1988). Constructing an acquisition-based procedure for second language assessment. *Studies in second language acquisition*, *10*(2), 217-243.

Pinker, S. (2013). *Learnability and Cognition: The Acquisition of Argument Structure*. MIT Press.

Quoc, V. & Schuster, M. (2016). *A Neural Network for Machine Translation, at Production Scale*. Retrieved from https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html

Radford, A. (2004). *English syntax: An introduction*. Cambridge University Press.

Rooth, M. (2004). *COMS474*: *Tree Syntax of Natural Language*, *Lecture note 1* [Lecture notes]. Retrieved from http://www.cs.cornell.edu/courses/cs474/2004fa/lec1.pdf

Schumann, J. (1979). The acquisition of English negation by speakers of Spanish: A review of the literature. *The acquisition and use of Spanish and English as first and second languages*, 3-32.

Siddharthan, A. (2002). An architecture for a text simplification system. In *Language Engineering Conference, 2002. Proceedings* (pp. 64-71). IEEE.

Siddharthan, A. (2006). Syntactic simplification and text cohesion. *Research on Language and Computation*, *4*(1), 77-109.

Simple English Wikipedia. (2019). *Wikipedia, The Free Encyclopedia*. Retrieved 10:28, June 18, 2019 from https://simple.wikipedia.org/w/index.php?title=Simple_English_Wikipedia&oldid=6576967.

Statt, N. (2016). *Google's AI translation system is approaching human-level accuracy*. Retrieved from https://www.theverge.com/2016/9/27/13078138/google-translate-ai-machine-learning-gnmt

Tillmann, C., & Ney, H. (2003). Word reordering and a dynamic programming beam search algorithm for statistical machine translation. *Computational linguistics*, *29*(1), 97-133.

Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive linguistics*, *11*(1/2), 61-82.

Velleman, E., & van der Geest, T. (2014). Online test tool to determine the CEFR reading comprehension level of text. *Procedia computer science*, *27*, 350-358.

Wang, T., Chen, P., Rochford, J., & Qiang, J. (2016). Text simplification using neural machine translation. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Wang, C., Collins, M., & Koehn, P. (2007). Chinese syntactic reordering for statistical machine translation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*.

Watanabe, W. M., Junior, A. C., Uzêda, V. R., Fortes, R. P. D. M., Pardo, T. A. S., & Aluísio, S. M. (2009). Facilita: reading assistance for low-literacy readers. In *Proceedings of* *the 27th ACM international conference on Design of communication* (pp. 29-36). ACM.

White, J. (1998). Getting the learners' attention: A typographical input enhancement study. *Focus on form in classroom second language acquisition*, 85-113.

White, J. (2008). Speeding up acquisition of his and her: Explicit L1/L2 contrasts help. *Second language acquisition and the younger learner: Child's play*, 193-228.

Wode, H. (1978). Developmental sequences in naturalistic L2 acquisition. *Second Language Acquisition: A Book of Readings*, 10-17

Wubben, S., Van Den Bosch, A., & Krahmer, E. (2012). Sentence simplification by monolingual machine translation. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1* (pp. 1015-1024). Association for Computational Linguistics.

Xiong, D., Liu, Q., & Lin, S. (2006, July). Maximum entropy based phrase reordering model for statistical machine translation. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics* (pp. 521-528). Association for Computational Linguistics.

Yamada, K., & Knight, K. (2001). A syntax-based statistical translation model. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*.

Zens, R., & Ney, H. (2003, July). A comparative study on reordering constraints in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1* (pp. 144-151). Association for Computational Linguistics.

Zhu, Z., Bernhard, D., & Gurevych, I. (2010). A monolingual tree-based translation model for sentence simplification. In *Proceedings of the 23rd international conference on computational linguistics* (pp. 1353-1361). Association for Computational Linguistics.

Zobl, H. (1982). A direction for contrastive analysis: The comparative study of developmental sequences. *TESOL quarterly*, *16*(2), 169-183.