Bachelor's Thesis

Submitted in Partial Fulfillment of the Requirements
for the Degree of
Bachelor of Arts in Computational Linguistics

Exploring Textual Cohesion Characteristics for German Readability Classification

Author:
Sabrina Galasso

Supervisor:
Prof. Dr. Walt Detmar Meurers

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

August 2014

Antiplagiatserklärung

Name: Galasso
Vorname: Sabrina
Matrikel-Nummer: 3730351
Adresse: Salmstr. 16/1, 72768 Reutlingen

Hiermit versichere ich, die Arbeit mit dem Titel:

„Exploring textual cohesion characteristics for German readability classification“

im Rahmen der Lehrveranstaltung „Intelligent Computer-Assisted Language Learning: Connecting CL research and real-life learning tasks“

im Sommersemester 2014 bei Prof. Detmar Meurers

selbständig und nur mit den in der Arbeit angegebenen Hilfsmitteln verfasst zu haben. Mir ist bekannt, dass ich alle schriftlichen Arbeiten, die ich im Verlauf meines Studiums als Studien- oder Prüfungsleistung einreiche, selbständig verfassen muss. Zitate sowie der Gebrauch von fremden Quellen und Hilfsmitteln müssen nach den Regeln wissenschaftlicher Dokumentation von mir eindeutig gekennzeichnet werden. Ich darf fremde Texte oder Textpassagen (auch aus dem Internet) nicht als meine eigenen ausgeben.

Ein Verstoß gegen diese Grundregeln wissenschaftlichen Arbeitens gilt als Täuschungs- bzw. Betrugsversuch und zieht entsprechende Konsequenzen nach sich. In jedem Fall wird die Leistung mit „nicht ausreichend“ (5,0) bewertet. In besonders schwerwiegenden Fällen kann der Prüfungsausschuss den Kandidaten/die Kandidatin von der Erbringung weiterer Prüfungsleistungen ausschließen (vgl. § 12 Abs. 3 der Prüfungsordnung für die Magisterstudiengänge vom 11. und 25. September 1995 bzw. § 13 Abs. 3 der Prüfungsordnung für die kulturwissenschaftlichen Bachelor- und Masterstudiengänge vom 12.10.2006 und 23.11.2007).

English version: I hereby declare that this paper is the result of my own independent scholarly work. I have acknowledged all the other authors' ideas and referenced direct quotations from their work (in the form of books, articles, essays, dissertations, and on the internet). No material other than that listed has been used.

Tübingen, August 29, 2014

S. Galasso

Sabrina Galasso

Contents

Abstract

This thesis investigates different feature sets that can be used to assess the cohesion of German texts including aspects of discourse coherence. The features are used to classify texts in terms of difficulty in a binary way. We focus on feature sets that were presented within researches for English and French, but not yet considered for German. This includes features based on different types of referring expressions, referential features, features based on connectives and features based on syntactic transitions. The implemented features will be evaluated on a corpus of German magazine articles that either address children or adults. The evaluation includes the features that were replicated from a previous study on this corpus. It is shown that the readability classification based on a combination of the implemented features and the replicated features results in an accuracy of 89.8%. The best performing feature subset implemented within this work is based on types of referring expressions, where especially pronoun frequencies make important contributions.

1 Introduction

Assessing the difficulty of texts is an emerging field of research in linguistics and cognitive science. An increasing demand arises from educational publishers. On the one hand, they aim to find reading material that suits the competencies of students to make sure that they understand the text or to offer them the possibility to train their reading skills. On the other hand, they are interested in understanding how the structure or complexity of a student's essays influences the grade level, to improve automated essay grading. Another field of applications is personalized web search, where it is tried to find documents in the web based on the reading level of web users [Collins-Thompson et al., 2011]. Furthermore, features that are developed to classify texts in terms of readability can reveal information that can be used for other types of classifications. Louwerse et al. [2004] found out that there exists a variation in cohesion across written and spoken registers.

Whether a text seems easy or difficult to a reader depends on many different factors. On the one hand, the surface of the text, such as the number of sentences, and more deeper linguistic characteristics of the text can influence its comprehensibility. The concept of *cohesion* is based on such characteristics. On the other hand, the reader's competencies regarding reading ability and world or domain knowledge need to be taken into account [McNamara et al., 2002]. This interaction among the text and the reader is considered when *coherence* is analysed. Within this thesis we will explore which characteristics of a text induce higher cohesion, which aims at classifying a text in terms of readability.

First, a background section will give an overview of existing approaches to readability assessment and discourse representation where various indices describe different characteristics of a text. Furthermore, it will introduce an existing experimental setup for German. The third section describes the feature sets that were implemented within this work to extend the existing setup. Section 4 will first describe the experimental procedure used to generate and evaluate the features. Afterwards we will present the results of the evaluation which includes also replicated features from Hancke et al. [2012]. The last section will draw a conclusion and suggest ideas on future work.

2 Background

Assessing the difficulty of texts is a persistent challenge in computational linguistics and is traditionally based on considering the surface of a text. The area of natural language processing made significant progress in the last decades, so that new techniques could be adapted to the tasks of readability classification including aspects of deeper linguistic structures. The following will give an overview of traditional and contemporary approaches and focus on two recent research projects on cohesion in English [McNamara et al., 2014] and French [Todirascu et al., 2013] texts. Afterwards I will introduce a German experimental setup for readability classification, which was implemented within the work of Hancke et al. [2012].

2.1 Measuring Readability

Traditional formulas for readability classifications mainly consider the surface of texts such as the length of a sentence. The *Flesch-Kincaid Grade Level* and the *Flesch Reading Ease* [Kincaid et al., 1975], which consider the length of the sentences and the number of syllables per word, were mostly used in the last decades. Another formula which arose early is based on the idea that commonly well known words are easier to process in terms of readability [Dale and Chall, 1948]. They generated a list of 3000 well known words and used corresponding counts to determine a text's difficulty.

More recent studies criticize these measures, since they do not investigate any deeper linguistic structures. McNamara et al. [1996] state that there exist phenomena which decrease such grades but increase the cohesion of a text in the same time. They present an experiment where four versions of an English text were generated manually by manipulating the cohesion of one text. They inserted connectives and increased the number of overlapping words among sentences. These modifications resulted in a lower *Flesch-Kincaid Grade Level*, but increased cohesion.

A widespread tool for evaluation of English text and discourse is *Coh-Metrix*[1] [McNamara et al., 2014]. It is used to evaluate English texts in terms of cohesion in a fully automatic way. The tool computes

[1] A web tool and its documentation is provided at `http://www.cohmetrix.com/`.

a broad range of fine-grained indices which can be grouped according to their underlying theoretical constructs, such as referential cohesion, connectives, lexical diversity, word frequencies, latent semantic analysis or readability measures. Tow tables in the appendix section A (p.25) list all the features provided in *Coh-Metrix*. The features based on referential characteristics among sentences and those based on connectives were adapted to German within this work. The number of paragraphs per text is one of the descriptive *Coh-Metrix* features and was also adapted to German.

- **Referential Features.** Most of the referential features implemented in *Coh-Metrix* analyse how much the sentences within a document correlate with each other, by counting words that occur in two sentences. The referential indices differ from each other in terms of the overlaps' explicitness, so that one index only takes nouns into account, whereas another index also includes overlaps among pronouns. For each type of explicitness there is also a variation concerning the distance among overlaps. *Coh-Metrix* differentiates between local cohesion, where one sentence is compared to its preceding sentence, and global cohesion, where every possible sentence pair is checked for overlaps. A more sophisticated way is considering all the words that refer to the same entity, i.e. allowing also pronouns to overlap with a noun given that they refer to the same entity. They provide a measure based on pronoun anaphora resolution, an automatic way to generate entity chains. Investigating referential cohesion is motivated by the assumption that sentences which do not share overlapping concepts cause cohesive gaps, which can have a negative impact on the reading time [McNamara et al., 2014, p. 63]. However, they state that the performance of anaphora resolution systems is modest [McNamara et al., 2014, p.51].
- **Connectives.** Another feature set is based on lists of different types of connectives, depending on the kind of connective link that is expressed by the word. They differentiate between *causal*, *temporal*, *additive*, *contrastive*, *logical*, *positive* (e.g. *also*, *moreover*) and *negative* (e.g. *however*, *but*) connectives. The motivation behind the consideration of connectives lies in the assumption that "explicitly linking ideas at the clausal and sentential level" [McNamara et al., 2014, p.46] increases discourse

cohesion.

McNamara et al. [2006] evaluated some of the referential and connective based features on a corpus of texts that were classified binary in terms of *high* and *low* cohesion by former experiments. They investigated the referential features that are not based on anaphora resolution, but on overlapping words or lemmas considering also the parts of speech. They extended the local overlaps to consider not only 2 adjacent sentences but also 3 and 4 which lead to significant results for all of them. Furthermore they created a list of *positive causal* connectives. A connective feature that was generated based on this list showed significant differences between the two data sets: The number of *positive causal* connectives increased, the higher the level of cohesion was.

The second research we will present is described within Todirascu et al. [2013]. They developed 41 features, mostly analysing discourse structures, and measured how well those features predict the difficulty of French texts that are used for teaching French as a foreign language. These texts were scaled according to the European standard *CERF* [Verhelst et al., 2009], which defines 6 levels of foreign language proficiency. The features were divided into the following six feature groups: *Part of speech tag-based variables*, *lexical coherence* and *entity coherence measures*, features investigating *entity density* and certain *properties of reference chains*. Within this work we will consider the following two feature sets for German.

- **Part of speech tag-based variables.** This feature set considers occurrences of pronouns, determiners and proper names and is motivated by the assumption that making use of certain types of referring expressions influences the cohesive level of a text. They tested their implemented features on a manually annotated corpus and got significant results for the variables *personal pronouns per sentence*, *definite articles per text* and *ratio of proper names per text*. More difficult texts tend to contain less personal pronouns per sentence and more definite articles per text than easier ones. The present work will explore these features for German on an automatically preprocessed corpus.
- **Entity coherence.** According to Todirascu et al. [2013] all the expressions referring to the same entity build a reference chain.

These chains were also annotated manually, since they wanted to investigate if the insignificance of implemented features in other studies is caused by errors made within the automatic annotation process. The feature set focuses on the four syntactic functions an entity can accept within a sentence: *subject*, *object*, *other complement* or *not present in the sentence*. To assess the local coherence, they observed how the entities' syntactic function changed within two adjacent sentences. They found out that the following transitions were significant in their manually annotated test data:

- *subject* → *object* transitions seem to occur more frequently in harder texts
- *object* → *object* transitions seem to occur more frequently in easier texts
- *subject* → *subject* transitions seem to occur more frequently in easier texts [2]

2.2 An Experimental Setup

A German setup for coherence assessment was already implemented within Hancke et al. [2012], where some features were motivated by *Coh-Metrix* indices. This setup was used as an experimental base line for this thesis. They performed readability classification on a German corpus using traditional features, lexical features, syntactic features, morphological features and features based on a language model.

2.2.1 The GEO-GEOlino Corpus

Motivated by studies on English texts using the *Weekly Reader*[3], Hancke et al. [2012] generated a corpus of articles extracted from the German reportage magazine *GEO*[4] and the magazine's children editions *GEOlino*[5]. In this way, a corpus of *easy* and *difficult* reports was extracted from the web to investigate differences in aspects of cohesion. For each of the three fields of interest (*Human*, *Nature* and *Technology*) the number of documents in *GEO* and *GEOlino* was

[2] However, they state that this feature is rarely observed and therefore it is not clear if the predictability is similar for other data

[3] `www.weeklyreader.com`

[4] `http://www.geo.de/`

[5] `http://www.geo.de/GEOlino/`

equal to avoid biases when training the classifier. Information about the computational preprocessing of the corpus is provided in Hancke et al. [2012].

2.2.2 Features Explored

They arranged their implemented features into five groups.

The first group (Trad) is composed of three traditional readability measures that consider sentence length, the number of syllables per word and the number of characters per word. These features were chosen despite of the contemporary criticism, since they have been used as standard measures for many years.

The group of lexical features (Lex) consists of features that are typical for German and of features adapted from the work of Lu [2012], who implemented them to judge the proficiency of second language English learners. The 23 lexical features investigate different variations and ratios of lexical words occurring in a text.

The set of syntactic features (Syn) is based on constituency based parse trees. They focus on three syntactic units (*sentences*, *clauses* and *T-Units*) which are defined by Lu [2010] and on characteristics at the phrasal level.

An additional feature set includes language modeling features (Lm) that are either only word based or additionally mixed with part of speech information. For both, they trained a unigram, a bigram and a trigram perplexity model on the *easy* and on the *difficult* data of another corpus, resulting in 12 Lm features. The reason for training the models on other data was to ensure that the results can be generalized across corpora [Hancke et al., 2012].

The morphological features (Morph) are highly related to the complex morphological structure of German and can be subdivided into inflectional, derivational and compound features. For verbal inflections they considered mood, person and tense, for nominal inflections only case. The derivational features consider for each of 25 derivational suffixes the *Suffix-Token Ratio*, the *Suffix-Noun Ratio* and the *Suffix-Derived Noun Ratio*. For compounds, the number of words that form a compound and the number of compounding components represents a feature.

2.2.3 Results

To evaluate the computed features, Hancke et al. [2012] used *WEKA*[6], a tool that provides a broad range of machine learning algorithms for data mining tasks [Hall et al., 2009]. They chose the Sequential Optimization (SMO) algorithm to train binary classifiers on different subsets and of combinations of these subsets. The 95 Morph features performed the best with an accuracy of 85.4%, followed by the feature sets Trad and Lex, both showing up about 82% accuracy. The Syn and Lm features performed worse with an accuracy of about 77%. Combining all the features resulted in an accuracy of 89.7%. The results show that the traditional approach of assessing readability was improved by adding more sophisticated features that reveal deeper linguistic structures. However, the only stand-alone feature set that outperforms traditional measures was the set of Morph features. This experimental setup is extended within this thesis, where four additional feature sets were implemented. The following will demonstrate their theoretical examination.

3 Features for Cohesion Assessment

Within this thesis 41 features were implemented to classify texts in terms of cohesion. These features can be grouped into 4 feature sets: referring expression based features (PrepDet), referential features (Ref), features based on connectives (Conn) and features based on syntactic transitions (Tran). Additionally, the number of paragraphs was counted. The implementation of the features was motivated by previous researches on cohesion of English and French texts [McNamara et al., 2014, Todirascu et al., 2013] and adapted to German.

3.1 Features on Types of Referring Expressions (PrepDet)

The investigation of features based on parts of speech is presented within Todirascu et al. [2013] and Pitler and Nenkova [2008] and concerns frequencies of pronouns, determiners and proper names. According to Hasan [1976], making use of certain referring expressions

[6] `http://www.cs.waikato.ac.nz/ml/weka/`

can lead to higher cohesion, since they construct a cohesive link to a specific referent, which is sometimes introduced earlier in the text. This way, a definite article links the corresponding noun to another sentence containing the information that introduces the entity [Hasan, 1976]. Thus, definite articles can be thought of increasing textual cohesion. The *RFTagger* [Schmid and Laws, 2008] identifies the parts of speech for German in a more fine-grained way. The underlying tagset is based on the *Stuttgart Tübingen Tagset(STTS)* [Schiller et al., 1995] but includes also morphological information such as the plurality and gender for nouns or the tense for finite verbs. In contrast to *STTS*, it differentiates between definite and indefinite articles, which is necessary for the intended investigation. The following lists the features implemented for this feature set:

- ratio of pronouns to nouns
- avg. proportion of pronouns per sentence and per text
- avg. proportion of personal pronouns per sentence and per text
- avg. proportion of possessive pronouns per sentence and per text
- avg. proportion of definite articles per sentence and per text
- ratio of proper names per text

The first feature shows how often pronouns occur with respect to the frequency of nouns. The following features consider the average proportions of *pronouns* in general, *personal pronouns*, *possessive pronouns* and *definite articles* per text and per sentence. The average proportion of an item per sentence takes into account the sentence length. If a sentence of length 8 contains 2 pronouns and another sentence of length 4 contains 1 pronoun, then the pronoun proportion of both is the same. Thus, these features give the same weight to short and long sentences. An additional feature investigates the ratio of proper names per text.

3.2 Referential Features (Ref)

The following describes the referential features that were implemented for this work. To compute most of the referential indices we count how often two sentences have at least one item in common. Since most of the indices do not consider the length of a sentence pair or the number of overlapping words, these indices are binary. Only the measures based on *content word overlaps* take into account how many overlaps occur with respect to the length of a sentence pair. Reusing the description of features in *Coh-Metrix* for a German application caused some changes due to different morphological structures of the languages. The following describes the re-implemented referential cohesion indices.

1. **Noun overlap**: The noun overlap indices count sentence pairs that contain at least one overlapping noun. Coh-Metrix does not allow for different word forms among overlaps. Thus, for English texts the words need to match in terms of plurality. For German this constraint is stronger than for English because of a more complex morphological structure. The two words $Häuser_{NOM,Plural}[houses]$ and $Häusern_{DAT,Plural}[houses]$ do match in plurality, but differ in case. Therefore they are not considered as a noun overlap within these indices. The index *local noun overlap* computes the average number of sentences that have an exact noun overlap to the previous sentence. The index *global noun overlap* computes the average number of all possible sentence pairs that contain an overlapping noun.
2. **Argument overlap**: The argument overlap indices count sentence pairs that contain at least one argument overlap. McNamara et al. [2014] count overlapping nouns, where plurality does not need to match (*family*↔*families*), and exactly overlapping pronouns (*he* ↔ *he*, *their* ↔ *their*). In natural language processing a noun is mostly lemmatized, when the plurality does not need to match. A lemma in a linguistic sense is the dictionary form of a certain word form. For German nouns, a lemmatizer does not only lead to ignoring plurality, but also to removing information about the word's case. Therefore, not only the number but also the case of the given noun is ignored. Thus, we check for matching lemmas ($Häuser_{NOM,Plural}[houses]$ ↔ $Hauses_{GEN,Singular}[house]$). Similar to the noun overlaps, there exists an index for the *local argument overlaps* and one for the *global argument overlaps*.
3. **Stem overlap**: The indices for stem overlaps count sentences that contain a noun that overlaps with a content word's stem contained in another sentence. A list of *STTS* part of speech

tags, which refer to content words was adopted from Hancke [2013]. The list mainly includes modal and full verbs, nouns, adjectives and adverbs. This measure aims to relax the noun constraint, so that the noun *Läufer*$_{NN}$[*runner*] would match the verb *laufen*$_{VVINF}$[*to run*]. Problematic to the *stem overlap* feature is the fact that the stemming tool does not perform well for German. The *Tartarus Snowball Stemmer*[7] for German, which is based on an algorithm described within Porter [1980], was used to stem the lemmas instead of the tokens themselves. Also for the *stem overlap*, the global correspondent was implemented.

4. **Content word overlap**: This measure indicates the proportion of explicit content words that overlap between two sentences taking into account the total number of content words contained in the sentence pair. It is the only measure within the referential features that is not binary and therefore useful for investigations, where the lengths of the sentences need to be taken into account. *Coh-Metrix* compares exact word forms. However, since the German morphology is more complex, lemmas were compared within this thesis. For adjectives the comparative forms share the same lemma (*gut*$_{ADV}$[*well*] ↔ *besser*$_{ADV}$[*better*]). The global correspondent is the proportion of explicit content words that overlap between pairs of sentences which do not need to be adjacent.

For German a stem overlap does not occur often, due to more complex morphological structures. Considering the root of a word might be more helpful than comparing stems. As an example, the root of the noun *Häufigkeit*$_{NN}$[*frequency*] and the root of the adjective *häufig*[*frequent*] are the same (*hauf*), whereas their stems (*haufig* $\nleftrightarrow$ *hauf*) do not match. Productive compounding prevents matches between words that would have been matching in English, where compounds also exist as spaced forms (e.g. *football shoe* vs. *Fußballschuh*). Applying a compound splitter first might improve the results. Additionally, the lexical semantic word net for German *GermaNet* [Henrich and Hinrichs, 2010] could be used to compute how often two sentences share words from a certain word class, such as *Gefühl*[*emotion*], *Geschehen*[*event*], *Motiv*[*reason*], since semantic relatedness increases cohesion.

[7] `http://snowball.tartarus.org/algorithms/german/stemmer.html`

3.3 Features based on Connectives (Conn)

Connectives are used to connect concepts within a text. The types of connectives describe the relation between the connected concepts. Out of the subclasses suggested by McNamara et al. [2014], the following were chosen to investigate for German:

- causal connectives
e.g. *daher* [*therefore*], *weil* [*because*]
- logical connectives
consisting of *und* [*and*], *oder* [*or*] and *wenn ... dann* [*if ... then*]
- temporal connectives
e.g. *dann* [*then*], *danach* [*afterwards*]
- additive connectives
e.g. *außerdem* [*furthermore*], *und* [*and*]
- adversative connectives
e.g. *wohingegen* [*whereas*]
- all connectives[8]

The lists of the namend connectives are taken from Eisenberg et al. [2009], a standard reference for German grammar. Since there were no predefined lists for *positive* and *negative* connectives available, these two subclasses of connectives were not investigated. The *adversative* connectives include also the Eisenberg et al. [2009]'s list of *concessive* connectives, since both types express a contrast and therefore are closely related to each other. The same holds for conditional connectives, which are included in the list of *causal* connectives, since Eisenberg et al. [2009] consider them as *causal* connectives in the "broader sense"[Eisenberg et al., 2009, p.1085].

The computed features represent incidence scores, which point out how often a connective or a certain type of connectives occurs per 1000 words. For each group of connectives, the following procedure is employed for every sentence:

1. count all the multiword connectives which do not allow for a distance in between them (e.g. *ohne dass* [a subordinating conjunction meaning *without*])

[8] By *all connectives* the union of all the named subclasses is meant

2. count all the multiword connectives which allow for a distance in between them (e.g. *weder ... noch* [*neither ... nor*]), provided that it was not already counted by 1.
3. count all the single word connectives, provided that it was not already counted by 1 or 2

This procedure avoids that a connective is counted twice within a group of connectives, but still allows to count them multiple times throughout all the connective types. Multiword connectives, which span across sentence boundaries are not counted.

A limitation of these features concerns two types of ambiguity in the list of connectives. First, it is not guaranteed that a string contained in the list is a connective in the context it is found. The following shows that the word *plötzlich* [*suddenly/sudden*] can be interpreted as a connective in example 1, but not in example 2.

(1) **Plötzlich**$_{ADV}$ begann es zu regnen.
Suddenly it started to rain.

(2) Der **plötzliche**$_{ADJA}$ Reichtum ist gefährlich.
The **sudden** wealth is dangerous.

Thus, at least the part of speech of the connectives found in the context should be considered, so that adjectives are not taken into account. The implemented features for German only consider words that are tagged as conjunction, adverb, pronominal adverb or adposition, since these parts of speech are required to capture most of the intended readings in the list of connectives. Table 3.1 provides German example connectives for every included part of speech tag. Filtering by part of speech avoids that the adjective *plötzlich* [*sudden*] in example 2 is counted as a connective. However, there are also cases where the part of speech is not sufficient to distinguish between connectives and non-connectives, as in the following examples:

(3) Die Erklärung war sehr gut. **So**$_{ADV}$ begann er zu verstehen, wie es funktionieren sollte.
The explanation was very clear. **So** he started to understand how it should work.

(4) Es ist **so**$_{ADV}$ kalt draußen.
It is **so** cold outside.

STTS **Tag**	**Description**	**Example Connective**
KON	coordinating conjunction	*jedoch*, *denn*
KOUS	subordinating conjunction	*nachdem*, *da*
KOUI	subordinating conjunction with infinitive construction	*anstatt (zu)*, *ohne (zu)*
APP*	preposition or postposition	*trotz*, *dank*
ADV	adverb	*bislang*, *später*
PAV	pronominal adverb	*trotzdem*, *deswegen*

Table 3.1: This table lists the STTS part of speech tags that can be assigned to connectives. The lists of connectives were taken from Eisenberg et al. [2009].

The second type of ambiguity refers to the ambiguity existing between subclasses of connectives. Some connectives are included in multiple subclasses, but the sense within a certain context is mostly restricted to one subclass. The connective *wenn* can be used as a temporal [*when*] or a causal [*if*] connective. In some context, even humans cannot distinguish among the classes.

A more sophisticated automatic approach to handle ambiguities based on part of speech tags would require to assign a tag to each connective by investigating all the possible meanings of a connective in the corresponding subclass. This would enable the investigation of connectives with a part of speech that is normally not considered as being appropriate for connectives. This holds especially for connectives that consist of multiple tokens, as illustrated in example 5. Within the implementation for this work, a past participle (*VVPP*) is not considered a connective and therefore it is not recognized. However, this procedure would require a fine-grained alignment of the STTS tags and the parts of speech assigned to connectives by Eisenberg et al. [2009] and would exceed the scope of this work. Furthermore it is not known if this affects the results concerning cohesion.

(5) **Abgesehen**$_{VVPP}$ **von**$_{APPR}$ dem schlechten Wetter hatten wir einen schönen Urlaub.
Apart from the bad weather we had a nice vacation.

3.4 Features based on Syntactic Transitions (Tran)

Todirascu et al. [2013] investigate the syntactic functions of entities occurring in a text. They manually annotated the expressions that refer to the same entity and observed the entity's syntactic function. Afterwards they analysed how the function changes throughout two adjacent sentences. They refer to Pitler and Nenkova [2008] who calculated the relative frequency of possible transitions between syntactic functions based on an entity coherence model for English texts [Barzilay and Lapata, 2008]. First, it will be shown how entities and syntactic functions can be extracted from dependency trees. Afterwards, the mentioned entity coherence model will be introduced to illustrate how the probabilities were calculated.

3.4.1 Extraction of Entities and Syntactic Functions

Within this thesis the dependency parser Bohnet and Kuhn [2012] was used to extract both, the entities themselves and their syntactic functions. The German parsing model is described in Seeker and Kuhn [2012]. Following Pitler and Nenkova [2008], we consider all noun phrases which share the same head noun as referring to the same entity. According to this theory, the noun phrase in example 6 refers to the same entity as the noun phrase in sentence 7. Therefore they represent one entity and the syntactic function of that entity changes from *subject* in example 6 to *object* in example 7.

(6) [Der kleine **Junge**]$_{NP,Subject}$ malt schöne Bilder.
[The small **boy**]$_{NP,Subject}$ paints nice pictures.

(7) Die Eltern loben [den fleißigen **Jungen**]$_{NP,Object}$.
The parents praise [the hardworking **boy**]$_{NP,Object}$.

Barzilay and Lapata [2008] define four possible functions that can be performed by an entity in a sentence: *subject*(S), *object*(O), *other complement*(X) or *not present in the sentence*(N). These functions were adapted to German and extracted from dependency parse trees. Table 3.2 lists, which tree labels were considered to belong to one of the caterogries S, O or X. The union of the three sets represents all the entities found in the text. It can later be used to generate for every sentence the fourth set N (*not occurring in the sentence*). As described before, only the nouns that represent the head of a noun phrase can represent an entity. It is checked for every occurring noun in a sentence, whether it falls into one of the three function categories, by considering its dependency label and in two cases also the labels of other dependent tokens. Nouns labeled with *NK* (*noun kernel modifier*) can be part of any of the three functions. Therefore the head needs to be considered, as can be seen in figure 3.1. The noun *Sandy* is labeled as *NK* with a preposition as its head. If the preposition is labeled as a prepositional object (OP), the noun is considered an object. In all the other cases (e.g. if the preposition is marked as a modifier) it is counted as *other complements*. Furthermore it is necessary to treat conjuncts in a special way: If the noun under consideration is a conjunct (*CJ*), the label of its closest dependent token determines the noun's function. To get the function *subject* for the entity *Lucy* in example 3.1 the label of *John* (*SB*) needs to be considered.

subjects	- subjects (SB) - noun kernel modifiers (NK): *consider the head* - conjuncts (CJ): *consider the closest head that is not a conjunct or coordinating conjunction*
objects	- accusative objects (OA,OA2) - dative objects (DA) - genitive objects (OG) - prepositional objects (OP) - noun kernel modifiers (NK): *consider the head* - conjuncts (CJ): *consider the closest head that is not a conjunct or coordinating conjunction*
other complements	- genitive attributes (AG) - parentheses (insertions) (PAR) - appositions (APP) - noun kernel modifier (NK): *consider the head* - conjuncts (CJ): *consider the closest head that is not a conjunct or coordinating conjunction*

Table 3.2: For each of the three functions *subject*, *object*, and *other complement* a set of entities can be extracted from dependency trees. This is done by considering the label of every noun occurring in the tree. The table lists for every function the possible dependency labels.

3.4.2 Feature Calculation

To illustrate how the features were calculated within this work, the following will introduce the term *Entity Grid* [Barzilay and Lapata, 2008]. An *Entity Grid* illustrates the information about entities and their functions in a structured way. Serving as example, we assume that the preceding sentences (6 and 7) were the first two sentences of a text containing *i* sentences and the noun *boy* would correspond

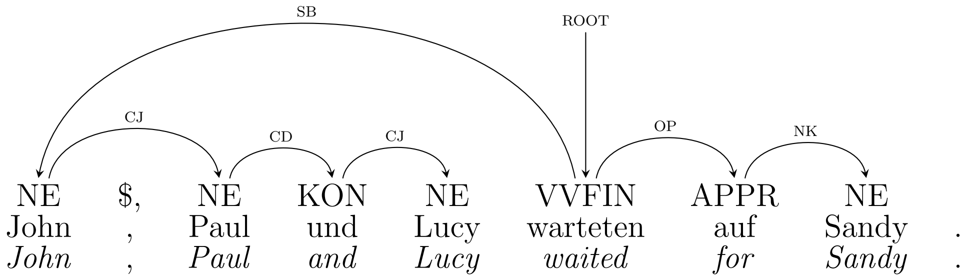
SB
ROOT
CJ
CD
CJ
OP
NK
NE $, NE KON NE VVFIN APPR NE
John , Paul und Lucy warteten auf Sandy .
John , Paul and Lucy waited for Sandy .

Figure 3.1: This figure visualizes the output of the dependency parser, which was used to extract entities and functions.

to *Entity 1*, with a total number of n entities throughout the whole text. This leads to an exemplary *Entity Grid* represented in table 3.3, where a transition from S to O is given for *Entity 1* from *Sentence 1* to *Sentence 2*.

	Entity 1	**Entity 2**	**Entity 3**	...	**Entity *n***
Sentence 1	S	O	N	...	N
Sentence 2	O	X	N	...	S
Sentence 3	S	N	N	...	N
Sentence ...	N	N	S	...	N
Sentence *i*	O	N	X	...	N

Table 3.3: For every sentence, a syntactic function is assigned to all the entities that occur in the text (including *N* meaning *not present in the sentence*). n is the number of entities in the text and i the number of sentences.

The features represent the average probability of each possible transition occurring in a text. The probability of a certain transition in a text is calculated as follows, where F_1 and F_2 refer to two syntactic functions (e.g. S and O), i is the number of sentences, and n the number of entities:

$$\Pr(F_1 \rightarrow F_2) = \frac{Count(F_1 \rightarrow F_2)}{(i-1) * n}$$

4 Experiments and Results

All the classification experiments were performed on the *GEO-GEOlino* corpus [Hancke et al., 2012] which was introduced in section 2.2.1. Their preprocessing procedure was modified, since new versions of the preprocessing tools were available. Due to technical problems during the calculation of the new features, some files had to be excluded [9]. To conduct a meaningful analysis and comparison to their implementation results, all their available feature calculations and evaluations were replicated using the same preprocessing procedure. The only feature that was computed on another version of the corpus was the *number of paragraphs*, since line breaks were not included in the current corpus version.

Following Hancke et al. [2012], all the classification models were created by employing the Sequential Minimal Optimization (SMO) algorithm, which is available with the machine learning tool *WEKA* [Hall et al., 2009]. To assess the quality of the classifier, we use 10-fold cross validation and present the overall accuracy [Hancke et al., 2012].

The following will first present the most predictive features implemented within this thesis. Then we will consider the results of the feature groups and further investigate their combinations, also including the features adapted from Hancke et al. [2012].

4.1 Most Predictive Features

WEKA [Hall et al., 2009] provides an algorithm for Information Gain, which extracts the most predictive features. Table 4.1 lists the ten most predictive features, which all together yield an accuracy of 76.1%. The 5 best features belong to the set which is based on types of referring expressions (PrepDet). Remarkably, these five features do all investigate pronouns. The features that explore determiners or proper names performed worse and are not part of the top 10 features. The local content word overlap is the only referential features in the given list. In contrast to all the other local overlap features, this measure takes into account the length of the sentences that share the overlapping word[10]. Furthermore the average probabilities of the two transitions $N \rightarrow N$ and $N \rightarrow O$ are part of the best performing features. They might be more reliable since they occur more often than transitions that do not contain an *N*. Another feature that predicts readability relatively well in combination with all the other features is the number of paragraphs in the text.

[9] From both sub-corpora (*GEO* and *GEOlino*) 4 files were excluded, keeping the number of files per genre equal

[10] The number of content words throughout the sentence pair is considered as the length of the sentence.

Feature	Feature Set
avg. proportion of personal pronouns per text	PrepDet
avg. proportion of pronouns per text	PrepDet
ratio of pronouns to nouns	PrepDet
avg. proportion of personal pronouns per sentence	PrepDet
avg. proportion of pronouns per sentence	PrepDet
local content word overlap	Ref
num of Paragraphs	-
avg. probability of $N \rightarrow N$ transition	Tran
incidence of causal connectives	Conn
avg. probability of $N \rightarrow O$ transition	Tran

Table 4.1: According to Information Gain, these are the ten most predictive features.

4.2 Feature Groups and Combinations

To evaluate the performance of the feature groups separately, a classifier was trained on various feature subsets. Table 4.2 shows for each of the four implemented feature groups how well they performed in terms of overall accuracy. The features based on types of referring expressions (PrepDet) outperform the other feature groups investigated within this thesis, with an accuracy of 74.6%. The second best performing set is based on connectives (Conn, 65.4%), closely followed by the referential features considering overlaps between sentences (Ref, 62.8%). Remarkably low is the accuracy of the features based on syntactic transitions (Tran, 51.4%), which suggests that *GEO* and *GEOlino* do not show differences when considering only this feature set. This result strengthens the assumption of Todirascu et al. [2013], who state that the significant results within their study might be achieved due to manual annotations. However, another reason for the different results of the studies might lay in the underlying data [Todirascu et al., 2013]. Moreover, the transitions that showed significance within their study might have rarely been observed within this thesis. The paragraph count per text (50.8%) does not seem to differ among the datasets, but it slightly improved the result for the total feature set (All) by approx 0.8%. This lead to an accuracy of 77.9% with the total set of 41 features. Classifiers were also trained on all possible combinations of the four subsets. The best performing combination (77.0%) contains the PrepDet, Conn and Ref feature groups, followed by the same set excluding the referential features

(76.5%). This shows that the PrepDet features do not only perform well independently as can be seen in the most predictive features, but also as a feature group.

Feature Set	Num. Features	Accuracy
PrepDet	10	74.6%
Ref	8	62.8%
Conn	6	65.4%
Tran	16	51.4%
number of Paragraphs	1	50.8%
PrepDet & Conn & Ref	24	77.0%
PrepDet & Conn	16	76.5%
All	41	77.9%

Table 4.2: The accuracy of all the feature sets implemented within this thesis and of several set combinations.

4.3 Combinations with Replicated Features

Almost all the features presented in Hancke et al. [2012] could be replicated.[11] Due to time issues, the Language Modeling features and some other individual features could not be replicated. This resulted in a set of 137 replicated features (instead of 155) with an accuracy of 88.7% (instead of 89.7%). Thus, the corresponding evaluation results listed in table 4.3 do not show remarkable differences towards the original feature results. Remarkably, the best result gained within this work (89.8%) was not achieved by putting together all the replicated and new developed features (89.5%), but by excluding those based on syntactic transitions (Tran).

Since the traditional features *sentence length*, *word length* and *syllables per word* are often used as a baseline in readability assessment, a classifier was trained on a combination of them (Trad) and all the features implemented within this thesis (All Thesis), which lead to an accuracy of 84.4%. This result shows for our dataset that adding the cohesive information described within this thesis to traditional readability measures improves the classification of *GEO* and *GEOlino* texts by approximately 3%.

Information Gain was also used to evaluate the whole resulting feature set. The ten most predictive features within the total set of

[11] Section 2.2.3 presents their results.

178 features include Lex features, Morph features, Syn features and 4 of the PrepDet features that were also among the top 10 features implemented within this thesis (Table 4.1 on page 18). Training a classifier on the ten most predictive features results in an accuracy of 84.6%.

Feature Set	Num. Features	Accuracy
Trad	3	81.2%
Lex	23	81.2%
Syn	24	74.6%
Morph	90	85.0%
All Replicated	137	88.7%
All Thesis & All Replicated	178	89.5%
All Thesis(excluding Tran) & All Replicated	163	89.8%
All Thesis & Trad	44	84.4%

Table 4.3: The accuracy of replicated features from Hancke [2013] and several combinations with features implemented within this thesis.

5 Conclusion

The underlying work gives an overview of approaches to readability classification for German by focusing on studies aiming at classifying English and French texts. Motivated by these studies, four feature sets were adapted to German: features based on types of referring expressions, referential features, features based on lists of connectives and features based on syntactic transitions.

The empirical baseline was adopted from Hancke et al. [2012], who generated a corpus containing an equal number of *easy* and *difficult* texts and implemented lexical, morphological, syntactic and language modeling features. The features that could be replicated from their work and the features implemented within this thesis were then evaluated following the methods applied in Hancke et al. [2012]. This procedure ensures that the results of their work and the results of this thesis are comparable. Out of the four new feature sets, the features based on referring expression performed best. This conclusion is drawn from the two applied evaluation methods. *Information Gain* was used to extract the ten most predictive features out of the 41 features described within this work. Five of these features belong to the group of PrepDet features, and are all based on pronoun frequencies. The second evaluation approach is the consideration of feature subsets, where the group of PrepDet features (74.6%) outperforms the second best feature group (Conn) by 9.2%. However, none of the feature sets outperforms independently the replicated traditional measures, which often serve as a baseline for readability classification.

Furthermore it was analysed how the implemented features perform in combination with the replicated features. The best result (89.8%) was achieved by excluding the features based on syntactic transitions. This features set performs worse as a stand-alone group, with an accuracy of 51.4%. Within Todirascu et al. [2013], this feature set was based on manual annotations to avoid errors caused by automatic annotations. Since this feature set is based on the investigation of entities, an automatic process of anaphora resolution might improve the results. The approach described in this work does not consider any pronouns. However, pronouns are crucial elements when it comes to the observation of entities. Thus, this feature set should either be evaluated on manual annotated data or its implementation should be based on a more complex entity recognition approach.

For future work, those features that could not be replicated from Hancke et al. [2012], especially the language modeling features, should be integrated into the evaluation. In addition it should be considered, how often the investigated items were actually observed. Certain syntactic transitions might occur so rarely that the results are not insightful enough to judge about their significance.

6 Acknowledgements

I would like to thank my supervisor Detmar Meurers for guiding me through this work. His persistent support and encouragement were an enormous help to me. Furthermore, I am grateful for the assistance given by Sowmya Vajjala and Tobias Kolditz. Their advice and comments have been a great help in implementing and evaluating my work. Finally, I would like to thank my family and friends for their valuable support throughout my study.

Bibliography

Regina Barzilay and Mirella Lapata. Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34, 2008.

Bernd Bohnet and Jonas Kuhn. The best of bothworlds – a graph-based completion model for transition-based parsers. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 77–87, Avignon, France, April 2012. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/E12-1009`.

Kevyn Collins-Thompson, Paul N Bennett, Ryen W White, Sebastian de la Chica, and David Sontag. Personalizing web search results by reading level. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 403–412. ACM, 2011.

Edgar Dale and Jeanne S Chall. A formula for predicting readability: Instructions. *Educational research bulletin*, pages 37–54, 1948.

Peter Eisenberg, Jörg Peters, Peter Gallmann, Catherine Fabricius-Hansen, Damaris Nübling, Irmhild Barz, Thomas A Fritz, and Reinhard Fiehler. *Duden 04. Die Grammatik - Unentbehrlich für richtiges Deutsch.* Band 4. Bibliographisches Institut GmbH, 8th edition, 2009.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

Julia Hancke. Automatic prediction of cefr proficiency levels based on linguistic features of learner language. Master's thesis, 2013.

Julia Hancke, Sowmya Vajjala, and Detmar Meurers. Readability classification for german using lexical, syntactic, and morphological features. In *Proceedings of COLING 2012: Technical Papers*, pages 1063–1080, 2012.

Halliday Hasan. *Cohesion in English.* Longman, London, 1976.

Verena Henrich and Erhard W Hinrichs. Gernedit-the germanet editing tool. In *ACL (System Demonstrations)*, pages 19–24, 2010.

Iuliia Ichin-Norbu. Assessing readability of german texts using discourse features. Unpublished Paper. Department of Linguistics, University of Tübingen, 2012.

J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, DTIC Document, 1975.

Max M Louwerse, Philip M McCarthy, Danielle S McNamara, and Arthur C Graesser. Variation in language and cohesion across written and spoken registers. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, pages 843–848, 2004.

Xiaofei Lu. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15 (4):474–496, 2010.

Xiaofei Lu. The relationship of lexical richness to the quality of esl learners' oral narratives. *The Modern Language Journal*, 96(2): 190–208, 2012.

Danielle S McNamara, Eileen Kintsch, Nancy Butler Songer, and Walter Kintsch. Are good texts always better? interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and instruction*, 14(1):1–43, 1996.

Danielle S McNamara, Max M Louwerse, and Arthur C Graesser. Coh-metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension. *Unpublished Grant proposal, University of Memphis, Memphis, Tennessee*, 2002.

Danielle S McNamara, Yasuhiro Ozuru, Arthur C Graesser, and Max Louwerse. Validating coh-metrix. In *Proceedings of the 28th annual conference of the cognitive science society*, pages 573–578. Erlbaum Mahwah, NJ, 2006.

Danielle S McNamara, Arthur C Graesser, Philip M McCarthy, and Zhiqiang Cai. *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press, 2014.

Emily Pitler and Ani Nenkova. Revisiting readability: A unified framework for predicting text quality. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 186–195. Association for Computational Linguistics, 2008.

Martin F Porter. An algorithm for suffix stripping. *Program: electronic library and information systems*, 14(3):130–137, 1980.

Anne Schiller, Simone Teufel, and Christine Thielen. Guidelines für das tagging deutscher textcorpora mit stts. *Universitäten Stuttgart und Tübingen*, 1995.

Helmut Schmid and Florian Laws. Estimation of conditional probabilities with decision trees and an application to fine-grained pos tagging. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 777–784. Association for Computational Linguistics, 2008.

Wolfgang Seeker and Jonas Kuhn. Making ellipses explicit in dependency conversion for a german treebank. In *LREC*, pages 3132–3139, 2012.

Amalia Todirascu, Thomas François, Nuria Gala, Cédrick Fairon, Anne-Laure Ligozat, and Delphine Bernhard. Coherence and cohesion for the assessment of text readability. *Natural Language Processing and Cognitive Science*, page 11, 2013.

N. Verhelst, Piet Van Avermaet, S. Takala, N. Figueras, and B. North. *Common European Framework of Reference for Languages: learning, teaching, assessment.* Cambridge University Press, 2009. ISBN 0521005310. URL `http://www.coe.int/T/DG4/Linguistic/CADRE_EN.asp`.

A An Overview of Existing Feature Implementations

The following appendix tables list for all the features described within [McNamara et al., 2014] and [Todirascu et al., 2013] if an equivalent is found in the German implementation setup. Comment *a* shows that the feature or a similar feature was implemented within Ichin-Norbu [2012]. Comment *b* means that the feature or a similar feature was implemented within Hancke [2013].

These are the features described within the study of Todirascu et al. [2013]:

Feature Set	**Variable**	**Feature**	**equivalent for German**	**comment**
POS tag-based variables	1	ratio between pronouns and nouns	B.A.Thesis	
	2	avg. proportion of pronouns per sentence	B.A.Thesis	
	3	avg. proportion of pronouns per word	B.A.Thesis	
	4	avg. proportion of personal pronouns per sentence	B.A.Thesis	
	5	avg. proportion of personal pronouns per word	B.A.Thesis	
	6	avg. proportion of possessive pronouns per sentence	B.A.Thesis	
	7	avg. proportion of possessive pronouns per word	B.A.Thesis	
	8	avg. proportion of definite articles per sentence	B.A.Thesis	
	9	avg. proportion of definite articles per sentence	B.A.Thesis	
	10	ratio of proper names per word	B.A.Thesis	
Lexical coherence measures	11	avg. similarity between adjacent sentences projected in a LSA space	-	
	12	word overlap (number of words in two consecutive sentences	-	
	13	lemma overlap		
	14	noun and pronouns overlap based on their lemmas	B.A.Thesis	
	15	noun and pronouns overlap based on their inflected forms	-	
Entity coherence	16-28	Relative frequency of the possible transitions between the four syntactic functions played by the entity in sentence n+1: subject (S), object (O), other complements (C), and (N) when the entity is absent .	B.A.Thesis	
Entity density	29	avg. proportion of entities per document	-	
	30	avg. number of entities per sentences	-	
	31	avg. proportion of unique entities per document	-	
	32	avg. number of words per entity	-	
reference chains	33	proportion of indefinite Nps included in a reference chain	-	
	34	proportion of definite Nps included in a reference chain	-	
	35	proportion of personal pronouns included in a reference chain	-	
	36	proportion of possessive determiners included in a reference chain	-	
	37	proportion of demonstrative determiners included in a reference chain	-	
	38	proportion of demonstrative pronouns included in a reference chain	-	
	39	proportion of reflexive pronouns included in a reference chain	-	
	40	proportion of proper nouns included in a reference chain	-	
	41	avg. length of reference chains	-	

These are the *Coh-Metrix* features 1 to 41:

Feature Set	**Variable**	**Feature**	**equivalent for German**	**comment**
Descriptive	1	Paragraph count, number of paragraphs	B.A.Thesis	
	2	Sentence count, number of sentences	found	
	3	Word count, number of words	-	
	4	Paragraph length, number of sentences, mean	B.A.Thesis	
	5	Paragraph length, number of sentences, standard deviation	-	
	6	Sentence length, number of words, mean	found	
	7	Sentence length, number of words, standard deviation	-	
	8	Word length, number of syllables, mean	found	
	9	Word length, number of syllables, standard deviation	-	
	10	Word length, number of letters, mean	found	
	11	Word length, number of letters, standard deviation	-	
Text Easability Principal Component Scores	12 - 27	see [McNamara et al., 2014] for further information	-	
Referential Cohesion	28	Noun overlap, adjacent sentences, binary, mean	B.A.Thesis	
	29	Argument overlap, adjacent sentences, binary, mean	B.A.Thesis	
	30	Stem overlap, adjacent sentences, binary, mean	B.A.Thesis	
	31	Noun overlap, all sentences, binary, mean	B.A.Thesis	
	32	Argument overlap, all sentences, binary, mean	B.A.Thesis	
	33	Stem overlap, all sentences, binary, mean	B.A.Thesis	
	34	Content word overlap, adjacent sentences, proportional, mean	B.A.Thesis	
	35	Content word overlap, adjacent sentences, proportional, standard deviation	-	
	36	Content word overlap, all sentences, proportional, mean	B.A.Thesis	
	37	Content word overlap, all sentences, proportional, standard deviation	-	
	38	Anaphor overlap, adjacent sentences	-	
	39	Anaphor overlap, all sentences	-	
LSA	40	LSA overlap, adjacent sentences, mean	-	
	41	LSA overlap, adjacent sentences, standard deviation	-	
	42	LSA overlap, all sentences in paragraph, mean	-	
	43	LSA overlap, all sentences in paragraph, standard deviation	-	
	44	LSA overlap, adjacent paragraphs, mean	-	
	45	LSA overlap, adjacent paragraphs, standard deviation	-	
	46	LSA given/new, sentences, mean	-	
	47	LSA given/new, sentences, standard deviation	-	
Lexical Diversity	48	Lexical diversity, type-token ratio, content word lemmas	-	a
	49	Lexical diversity, type-token ratio, all words	found	
	50	Lexical diversity, MTLD, all words	-	b
	51	Lexical diversity, VOCD, all words	-	b
Connectives	52	All connectives incidence	B.A.Thesis	
	53	Causal connectives incidence	B.A.Thesis	
	54	Logical connectives incidence	B.A.Thesis	
	55	Adversative and contrastive connectives incidence	B.A.Thesis	
	56	Temporal connectives incidence	B.A.Thesis	
	57	Expanded temporal connectives incidence	-	
	58	Additive connectives incidence	B.A.Thesis	
	59	Positive connectives incidence	-	
	60	Negative connectives incidence	-	

These are the *Coh-Metrix* features 42 to 108:

Feature Set	Variable	Feature	equivalent for German	comment
Situation Model	61	Causal verb incidence	-	a
	62	Causal verbs and causal particles incidence	-	a
	63	Intentional verbs incidence	-	a
	64	Ratio of casual particles to causal verbs	-	a
	65	Ratio of intentional particles to intentional verbs	-	a
	66	LSA verb overlap	-	
	67	WordNet verb overlap	-	
	68	Temporal cohesion, tense and aspect repetition, mean	-	a
Syntactic Complexity	69	Left embeddedness, words before main verb, mean	-	
	70	Number of modifiers per noun phrase, mean	found	
	71	Minimal Edit Distance, part of speech	-	
	72	Minimal Edit Distance, all words	-	
	73	Minimal Edit Distance, lemmas	-	
	74	Sentence syntax similarity, adjacent sentences, mean.	-	
	75	Sentence syntax similarity, all combinations, across paragraphs, mean	-	
Syntactic Pattern Density	76	Noun phrase density, incidence	-	b
	77	Verb phrase density, incidence	-	b
	78	Adverbial phrase density, incidence	-	
	79	Preposition phrase density, incidence	-	b
	80	Agentless passive voice density, incidence	-	b
	81	Negation density, incidence	-	
	82	Gerund density, incidence	-	
	83	Infinitive density, incidence	-	
Word Information	84	Noun incidence	-	b
	85	Verb incidence	-	b
	86	Adjective incidence	-	
	87	Adverb incidence	-	
	88	Pronoun incidence	-	
	89	First person singular pronoun incidence	-	
	90	First person plural pronoun incidence	-	
	91	Second person pronoun incidence	-	
	92	Third person singular pronoun incidence	-	
	93	Third person plural pronoun incidence	-	
	94	CELEX word frequency for content words, mean	-	
	95	CELEX Log frequency for all words, mean	found	
	96	CELEX Log minimum frequency for content words, mean	-	
	97	Age of acquisition for content words, mean	-	
	98	Familiarity for content words, mean	-	
	99	Concreteness for content words, mean	-	
	100	Imagability for content words, mean	-	
	101	Meaningfulness, Colorado norms, content words, mean	-	
	102	Polysemy for content words, mean	found	
	103	Hypernymy for nouns, mean	-	b
	104	Hypernymy for verbs, mean	-	b
	105	Hypernymy for nouns and verbs, mean	-	b
Readability	106	Flesch Reading Ease	-	
	107	Flesch-Kincaid Grade Level	-	a
	108	Coh-Metrix L2 Readability	-	a