

Modeling the Readability of German Targeting Adults and Children: An empirically broad analysis and its cross-corpus validation

Zarah Weiß Detmar Meurers
ICALL-Research.com
Department of Linguistics
University of Tübingen
{zweiss, dm}@sfs.uni-tuebingen.de

Abstract

We analyze two novel data sets of German educational media texts targeting adults and children. The analysis is based on 400 automatically extracted measures of linguistic complexity from a wide range of linguistic domains. We show that both data sets exhibit broad linguistic adaptation to the target audience, which generalizes across both data sets. Our most successful binary classification model for German readability robustly shows high accuracy between 89.4%–98.9% for both data sets. To our knowledge, this comprehensive German readability model is the first for which robust cross-corpus performance has been shown. The research also contributes resources for German readability assessment that are externally validated as successful for different target audiences: we compiled a new corpus of German news broadcast subtitles, the *Tagesschau/Logo* corpus, and crawled a *GEO/GEolino* corpus substantially enlarging the data compiled by Hancke et al. (2012).

Zusammenfassung

Wir untersuchen zwei neue Datensätze deutscher Bildungs- und Mediensprache für Kinder und Erwachsene. Die Analyse basiert auf 400 automatisch extrahierten Maßen sprachlicher Komplexität, die verschiedene linguistische Domänen abdecken. Unsere Ergebnisse zeigen, dass in beiden Datensätzen die sprachliche Gestaltung der Texte in ähnlicher Weise breitflächig an ihr jeweiliges Zielpublikum angepasst wird. Unser erfolgreichstes binäres Klassifikationsmodell erzielt Genauigkeitswerte von 89,4% und 98,9% über beide Datensätze hinweg. Unseres Wissens handelt es sich bei diesem umfassend durch verschiedene linguistische Bereiche informiertem Modell deutscher Text-Lesbarkeit um das erste, für das robuste Ergebnisse in einer korpusübergreifenden Evaluation dokumentiert sind. Darüber hinaus tragen wir mit unserer Arbeit zwei neue Datensätze zur Erforschung deutscher Text-Lesbarkeit bei, die auf Texten basieren, deren Eignung für ihre respektiven Zielgruppen extern durch wiederholte Rezeption validiert wurde: Wir haben aus Untertiteln deutscher Nachrichtenbeiträge das *Tagesschau/Logo* Korpus erstellt. Weiterhin haben wir das *GEO/GEolino* Korpus beträchtlich erweitert, das ursprünglich von Hancke et al. (2012) erstellt wurde.

1 Introduction

Readability assessment refers to the task of (automatically) linking a text to the appropriate target audience based on its complexity. A diverse spectrum of potential application domains has been identified for this task in the literature, ranging from the design and evaluation of education materials, to information retrieval, and text simplification. Given the increasing need for learning material adapted to different audiences and the barrier-free access to information required for political and social participation, automatic readability assessment is of immediate social relevance. Accordingly, it has attracted considerable

This work is licensed under a Creative Commons Attribution 4.0 International License. License details: <http://creativecommons.org/licenses/by/4.0/>

research interest over the last decades, particularly for the assessment of English (Crossley et al., 2011; Chen and Meurers, 2017; Feng et al., 2010).

For German readability assessment, however, little progress has been made in recent years, despite a series of promising results published around the turn of the decade (Vor der Brück et al., 2008; Hancke et al., 2012). In particular, German readability research has suffered from the lack of a shared reference corpus and sufficiently comparable corpora for cross-corpus testing of readability models: While for English research, the *Common Core* corpus consisting of examples from the English Language Arts Standards of the Common Core State Standards, and the *WeeklyReader* corpus of online news articles have been widely used in studies on English readability and text simplification (Vajjala and Meurers, 2014; Petersen and Ostendorf, 2009; Feng et al., 2010), there are no comparable resources for German. This is particularly problematic, as over-fitting is a potential issue for classification algorithms, especially given the limited size of the typical data sets.

To address these issues, we first present two new data sets for German readability assessment in Section 3: a set of German news broadcast subtitles based on the primary German TV news outlet *Tagesschau* and the children’s counterpart *Logo!*, and a GEO/GEOlino corpus crawled from the educational GEO magazine’s web site, a source first identified by Hancke et al. (2012), but double in size.¹ The longstanding success of these outlets with their target audiences provides some external validity to the nature of the implicit linguistic adaptation of the language used. As Bryant et al. (2017) showed for German secondary school textbooks, this is not necessarily the case across all linguistic dimensions and adjustments may even be limited to only the surface level of text, sentence, and word length. We conducted a series of analyses on these two data sets to accomplish the following objectives:

1. Investigate how instances of German educational news language differ in terms of language complexity across adult and child target audiences.
2. Build a binary readability model for German educational language targeting adults and children that shows high, robust classification performance across corpora.

For the purposes of our studies, we operationalize child target audience of German educational news language as children aged between 8 and 14. This is the typical audience age range of the child-targeting news media we analyzed.² Adult target audience then is defined as over 14 years of age.

To address our first research question, after introducing a broad set of complexity measures in Section 4, we compare their informativeness for distinguishing adult and child level in the two data sets in Section 5. In Section 6, we define a series of readability models for German, including one showing high classification accuracy between 89.4% and 98.9% on both data sets. The paper closes with a discussion of the implications of our results for the current research discussion and an outlook on future work.

2 Related Work

For over a century, text readability has been assessed using surface measure-based readability formula such as the Flesch-Kincaid formula (Kincaid et al., 1975) or the Dale-Chall readability formula (Chall and Dale, 1995), see for an overview DuBay (2004). While these formula are still used in some non-linguistic studies (Woodmansey, 2010; Grootens-Wiegers et al., 2015; Esfahani et al., 2016), a decade ago research shifted towards using more elaborate statistical modeling approaches based on larger sets of linguistically more informed features. Automatic readability assessment has benefited from the use of Natural Language Processing tools for the assessment of syntactic, lexical, and discourse measures and from adapting complexity measures employed in Second Language Acquisition research (Vajjala and Meurers, 2012; Feng et al., 2010). There has also been extensive research on the relevance of cohesion

¹We are currently negotiating with the broadcasters of *Tagesschau* and *Logo!* and the publishers of *GEO/GEOlino* to make the data freely available to other researchers and will make it available from <http://www.icall-research.de> in that case.

²The *GEOlino* magazine is advertised as targeting children between 8 and 14 years (cf. <http://www.geo.de/magazine/geolino-magazin>, accessed 11.06.18, 15:49). *Logo!* does not specify the age of its target audience, but has been reported to be particularly popular with children from age 8 to 12 (vom Orde, 2015).

and discourse measures for readability assessment that have successfully been employed for proficiency assessment in the *CohMetrix* project (Crossley et al., 2008; Crossley et al., 2011). Another example is the work by Feng et al. (2010), who evaluate which of the typically proposed measures of text readability are most promising by studying their relevance on primary school students reading material. They find language model features and cohesion in terms of entity density to be particularly useful, as well as measures of nouns. Interestingly, they also observe overall sentence length to be more informative than more elaborate syntactic features. While Feng et al. (2010) do not elaborate further on other lexical measures than POS features, Chen and Meurers (2017) conduct an elaborate cross-corpus study on the use of word frequency features for readability assessment. They show, that the typical aggregation of word frequencies across documents are less informative than richer representations including frequency standard deviations.

In contrast to English, research on readability assessment for other languages, such as German, is more limited. There was a series of articles on this issue from the late 2000s to the early 2010s that demonstrated the benefits of broad linguistic modeling, in particular the use of morphological complexity measures for languages with rich morphological systems like German (Vor der Brück et al., 2008; Hancke et al., 2012), but also Russian (Reynolds, 2016) or French (François and Fairon, 2012). The readability checker *DeLite* of Vor der Brück et al. (2008) is one of the first more sophisticated approaches that went beyond using simple readability formulas for German. The tool employs morphological, lexical, syntactical, semantic, and discourse measures, which they trained on municipal administration texts rated for their readability by humans in an online readability study involving 500 texts and 300 participant, resulting in overall 3,000 ratings. However, due to the specific nature of the data, the robustness of the approach across genres is unclear. Municipal administration language is so particular that results are unlikely to generalize to educational or literary materials, which are more attractive in first and second language acquisition contexts.

Later work by Hancke et al. (2012) also combines traditional readability formula measures, such as text or word length, with more sophisticated lexical, syntactic, and language model, and morphological features to assess German readability, but they employ an overall broader and more diverse feature set than *DeLite*. They investigate readability of educational magazines on the GEO/GEOLino data set, which they compiled from online articles freely available at the GEO magazine's web page. Their work illustrates the relevance of rich linguistic modeling for readability assessment and in particular the value of morphological complexity features for German.

The latest large scale research endeavor for the assessment of German text readability has focused more on identifying linguistic differences between texts targeting different audiences than on building readability models: In the *Reading Demands* project, complexity differences in German secondary school book texts across grade levels and school types were investigated. Berendes et al. (2017) and Bryant et al. (2017) analyze to which extent publishers successfully adapt their reading material to their target audiences. They find a lack of consistent adaptation for passive constructions, concessive and adversative connectives, and relative clauses, and only some limited adaptation in terms of lexical variation, noun complexity, and dependency length measures.

3 Data Sets

3.1 GEO/GEOLino

The GEO/GEOLino data set consists of online articles from one of the leading German monthly educational magazines, *GEO*, and the counterpart for children, *GEOLino*.³ They are comparable to the *National Geographic* magazine and cover a variety of topics ranging from culture and history to technology and nature. Hancke et al. (2012) first compiled and analyzed a data set from this web resource. We followed their lead and crawled 8,263 articles from the GEO/GEOLino online archive, almost doubling the size of the original corpus. We removed all material flagged as non-article contents by GEO as well as all articles that contained less than 15 words. We further cleaned our data from crawling artifacts and performed near-duplicate detection with the Simhash algorithm. We then grouped all texts into topic categories

³<http://www.geo.de> and <http://www.geo.de/geolino>

based on the subdomains they were published under, following the web page topic structure.⁴ Table 1 shows the composition of the corpus in terms of the topic groups. Since the number of documents in the different topic groups differ between GEO and the smaller GEOLino set, we created a more balanced subset (GEO/GEOLino_S). For this, we included only topic categories existing in both GEO and GEOLino, included all GEOLino texts in those categories and sampled from the GEO texts in those categories until we reached the same overall size of 2480 texts each.

Topic	GEO	GEOLino	Σ	GEO _S	GEOLino _S	Σ_S
Do It Yourself	0	663	663	0	0	0
Humanity	1,476	1,168	2,644	1,047	1,168	2,215
Nature	1,704	576	2,280	1,218	576	1,794
Reviews	300	736	1,036	215	736	951
Technology	0	121	121	0	0	0
Travel	1,519	0	1,519	0	0	0
Σ	4,999	3,264	8,263	2,480	2,480	4,960

Table 1: Distribution of topics in the full and sampled GEO/GEOLino data set.

3.2 Tagesschau/Logo

The Tagesschau/Logo data set is compiled from subtitles of German daily news broadcasts of *Tagesschau* and its children’s counterpart *Logo!*. *Tagesschau* is the dominant national television news service of Germany, produced by the German public-service television network ARD. It broadcasts multiple updated editions of daily news throughout the day. *Logo!* is a television news service for children produced by the German public-service television broadcaster ZDF airing once a day. The data set consists of subtitles for all editions of both news outlets that have been broadcasted from December 2015 to January 2017. For this paper, we limited the *Tagesschau* data to the main edition broadcasted at 8pm. This amounts to overall 421 editions for *Tagesschau* and 415 editions for *Logo!*, with the small difference arising from a lack of *Logo!* broadcasts on some public holidays or due to special broadcasts. We cleaned the subtitles by removing non-spoken comments (e.g., ** music playing ** or ** cheering **).

3.3 Characteristics of the two data sets

Table 2 compares the profiles of the GEO/GEOLino_S and the Tagesschau/Logo data sets that we used.

	GEO _S	GEOLino _S	Tagesschau	Logo
Num. Documents (total)	2,480	2,480	421	415
Num. Words (median)	383	350	1631	1322
Num. Sentences (median)	23	25	167	125

Table 2: Corpus profile for sampled GEO/GEOLino data set and the Tagesschau/Logo data set.

While GEO/GEOLino contains more documents than Tagesschau/Logo, they are considerably shorter in terms of the number of words and sentences they contain. Another difference arises in terms of the medium: GEO/GEOLino articles are self-contained reading material and Tagesschau/Logo subtitles

⁴Subdomains were mapped to topic groups in the following way based on the URL components following <http://www.geo.de> and <http://www.geo.de/geolino>: building (“basteln”), learning (“lernen”), children’s recipes (“kinderrezepte”), and competitions (“wettbewerbe”) were categorized as DO IT YOURSELF. Jobs (“berufe”), extras (“extras”), photography (“fotografie”), creativity (“kreativ”), info (“info”), love (“liebe”), magazines (“magazine”), human (“mensch”), idioms (“redewendungen”), and knowledge (“wissen”) were categorized as HUMANITY. Nature (“natur”), nature and environment (“natur-und-umwelt”), and animal encyclopedia (“tierlexikon”) were labeled as NATURE. Book reviews (“buechertipps”), movie reviews (“filmtipps”), game reviews (“spieletest”), and GEO television (“geo-tv”) were labeled as REVIEWS. Research and technology (“forschung-und-technik”) was labeled as TECHNOLOGY. Travel (“reisen”) was labeled as TRAVEL.

complement video material. At the same time, they consist of German educational media language and share the functional goal of conveying information to the reader, so that we consider them to be sufficiently similar to support a cross-corpus analysis.

4 Complexity Analysis

For the assessment of German language complexity, we extract 400 complexity measures using state of the art NLP techniques. All features are theoretically grounded in the contemporary research in linguistic subdisciplines, in particular Second Language Acquisition research, where Complexity is one of three dimensions of language proficiency, together with Accuracy and Fluency (Housen et al., 2012). SLA research has a rich tradition of analyzing the complexity development of learner language, see Lu (2010; 2012) for an overview. Vajjala and Meurers (2012) show that these measures can be successfully applied to readability research. Building on these findings, we follow the SLA definition of complexity as the elaborateness and variability of language (Ellis and Barkhuizen, 2005). Our measures can be grouped into seven categories: i) lexical complexity, ii) clausal complexity, iii) phrasal complexity, iv) morphological complexity, v) discourse complexity, vi) cognitive complexity, and vii) language use. While the former five groups are rooted in the linguistic system, the latter two categories were derived from psycholinguistic research. The resulting complexity assessment covers a broad variety of measures. To the best of our knowledge, this is currently the most extensive feature collection for German complexity assessment.⁵ Table 3 gives an overview of the feature categories and how much they contribute to our assessment.⁶

Category	#	Description
Descriptive	2	Total number of sentences and words.
Lexical	73	Lexical diversity measures such as general and POS-specific type-token ratios as well as semantic relatedness measures.
Sentential	119	Ratios measuring sentential elaboration and variation, such as clauses per sentence.
Phrasal	41	Ratios measuring phrasal elaboration and variation, such as modifiers per noun phrase.
Morphological	39	Ratios of inflection, derivation, and composition measures.
Cohesion	48	Subsequent (local) or across text (global) use of implicit or explicit cohesion markers such as connectives, pronouns, or grammatical transitions.
Cognitive	23	Dependency lengths, verb-argument distances, and ratios of cognitive integration costs assessing cognitive processing load based on Gibson’s (2000) Dependency Locality Theory.
Language Use	54	Word frequency ratios based on Subtlex-DE (Brysbaert et al., 2011), dlexDB (Heister et al., 2011), Karlsruhe Children’s Texts (Lavalley et al., 2015) Approximation of age of active use based on Karlsruhe Children’s Texts.

Table 3: Overview over complexity measures grouped by feature categories.

In order to extract these measures, we employ an elaborate analysis pipeline which relies on a number of NLP tools and external linguistic resources. We use OpenNLP 1.6.0 for tokenization and sentence segmentation.⁷ This serves as input for the Mate tools 3.6.0 (Bohnet and Nivre, 2012), which perform a morphological analysis, lemmatization, POS tagging, and dependency parsing. We then use the JWord-

⁵Our feature collection draws from varying perspectives on language complexity including SLA and human language processing research. While the confirmation or refutation of specific theories underlying these measures is an interesting research endeavor, our empirical questions focus on which of these features support the distinction of texts targeting different audiences.

⁶We are working on integrating our German complexity analysis pipeline into CTAP (Chen and Meurers, 2016) to make it generally available and will include an online documentation for each feature.

⁷<http://opennlp.apache.org>

Splitter 3.4.0 for compound analysis.⁸ The Mate POS tags are further used to inform the Stanford PCFG parser 3.6.0 (Rafferty and Manning, 2008) and the Berkeley parser 1.7.0 (Petrov and Klein, 2007), which we use for constituency and topological field parsing. For all tools, we use the German default models that were provided with them, except for the Berkeley parser, for which we use the topological field model by Ziai and Meurers (2018). With these annotations, we extract all instances of the linguistic constructs that we need to calculate the final 400 complexity ratios.⁹

5 Study 1: Which complexity measures are informative?

5.1 Set-Up

We first want to determine the informativeness of each measure for distinguishing between adult and child target audience. For this, we calculate the information gain of each measure on both data sets using 10-folds cross-validation for training and testing. We then compare across both data sets i) the number of features that are informative, and ii) the 20 most informative measures that show a Pearson correlation smaller than ± 0.8 with each other.¹⁰ This allows us to gain insights into the range of linguistic properties of the documents targeting adults and children. We used WEKA (Hall et al., 2009) to calculate information gain and R for the correlation analysis.

5.2 Results and Discussion

Table 4 shows the percentage of measures that exhibited an average information gain above zero.

Data Set	Percentage	Informative to Total
GEO/GEOlino	79.00%	316/400
Tagesschau/Logo	88.25%	353/400

Table 4: Percentage of informative measures based on 10-folds cross-validated information gain.

Overall, 79.00% of the measures are informative for the GEO/GEOlino data and 88.25% for the Tagesschau/Logo data. This shows, that the documents are adjusted to their different target audiences in terms of a broad range of dimensions of linguistic complexity.

Table 5 provides a deeper look into the linguistic design of the documents by showing the 20 most informative measures distinguishing adult from child targeted documents, including only measures with a correlation less than ± 0.8 . The table shows the original rank of each measure before removal of correlated measures, the average merit of each measure for the distinction of the target audience, the type of complexity measures it belongs to, and the feature name.

The results for both data sets show a diverse collection of features, some of which are similar for both data sets, but also some interesting differences. In total the measures seem to be more informative for Tagesschau/Logo, as indicated by the higher average merit, and more correlated, as can be seen from the wider range of original ranks. Language use as captured by frequency measures is particularly relevant for both data sets. The table includes seven measures of word frequency for GEO/GEOlino and five for Tagesschau/Logo. For both data sets, the most informative measure is one of language use: For GEO/GEOlino it is the average minimal age of active use of lexical types found in the Karlsruhe Children’s Corpus (KCT) of Lavalley et al. (2015). For Tagesschau/Logo it is the average log lexical type frequency based on Google Books 2000. The other language-use measures are very similar across data sets: Lexical types unknown to the Subtlex-DE data base (Brysbaert et al., 2011), for example, rank 4th and 2nd on both data sets and while on Tagesschau/Logo the lemma frequency per lexical type found in KCT is the 12th most informative measure, its log counterpart ranks 8th on GEO/GEOlino.

⁸<http://www.danielnaber.de/jwordsplitter>

⁹To support transparent comparison with other complexity studies, we include a description of the operationalization of all linguistic units that allow for varying definitions in Appendix A, as has been suggested by Bulté and Housen (2014).

¹⁰We set the Pearson correlation threshold relatively high since we primarily are interested in qualitatively inspecting the types of measures that are informative, not in removing all correlations.

GEO/GEOlino				Tagesschau/Logo			
Rank	Average Merit	Group	Feature	Rank	Average Merit	Group	Feature
1	0.332 (± 0.004)	USE	sumTypesMinAoAPerTypeInKCT	1	0.978 (± 0.004)	USE	logTypeFreqsPerTypeInGoogle00
4	0.327 (± 0.005)	LEX	syllablesPerToken	31	0.899 (± 0.006)	USE	typesNotInSubtexPerLexicalType
11	0.288 (± 0.004)	USE	logTypeFreqsPerTypeInSubtex	50	0.825 (± 0.009)	COH	2PPersPronounsPerNoun
13	0.231 (± 0.003)	USE	typesNotInSubtexPerLexicalType	71	0.754 (± 0.012)	COH	probNotSubsPerTransition
15	0.205 (± 0.003)	COH	2PPersAndPossPronounsPerToken	82	0.716 (± 0.010)	COH	causalConnectivePerSentence
21	0.164 (± 0.004)	USE	typesNotInDlexPerLexicalType	85	0.689 (± 0.009)	COH	localArgOverlapsPerSentence
23	0.147 (± 0.004)	PHR	complexNominalsPerTUnit	88	0.667 (± 0.008)	SEN	sumParseTreeHeightsPerFiniteClause
24	0.143 (± 0.002)	USE	logLemmaFreqsPerTypeInKCT	89	0.662 (± 0.008)	SEN	NPsPerTUnit
25	0.143 (± 0.003)	SEN	syllablesInMiddleFieldPerMiddleField	90	0.656 (± 0.007)	COH	1PPersPronounsPerToken
28	0.133 (± 0.003)	COH	persPronounsPerToken	91	0.657 (± 0.010)	MOR	genitivesPerNoun
31	0.133 (± 0.003)	SEN	PPsPerTUnit	95	0.633 (± 0.011)	PHR	determinersPerNP
33	0.132 (± 0.003)	MOR	secondPersonMarkingsPerFiniteVerb	100	0.622 (± 0.011)	USE	lemmaFreqsPerTypeInKCT
35	0.123 (± 0.004)	MOR	ionTPerToken	101	0.620 (± 0.014)	COG	sumLongestDependenciesPerClause
36	0.122 (± 0.003)	LEX	synsetPerTypeInGnet	102	0.617 (± 0.008)	MOR	compundNounsPerNP
37	0.121 (± 0.004)	PHR	complexNominalsPerFiniteClause	103	0.609 (± 0.012)	USE	typeFreqsPerTypeInDlex
38	0.120 (± 0.004)	SEN	sumNonTerminalNodesPerTUnit	109	0.568 (± 0.008)	LEX	MTLD
40	0.118 (± 0.004)	USE	typeFreqsPerTypeInSubtex	110	0.560 (± 0.010)	LEX	nonAuxVerbTypesPerNonAuxVerbToken
43	0.114 (± 0.002)	COH	pronounsPerNoun	111	0.550 (± 0.013)	COH	globalStemOverlapsPerSentence
44	0.113 (± 0.002)	USE	logAnnoTypeFreqBand5PerTypeInKCT	117	0.505 (± 0.011)	SEN	conjunctioalClausesPerSentence
49	0.111 (± 0.002)	COH	3PPersAndPossPronounsPerNoun	119	0.500 (± 0.014)	USE	logAnnoTypeFreqBd4PerTypeInDlex

Table 5: Top 20 most informative measures on balanced GEO/GEOlino and Tagesschau/Logo data based on information gain with $r \leq 0.8$.

Cohesion measures are highly informative, too, although more so for Tagesschau/Logo. In particular the use of certain personal or possessive pronouns is highly informative for GEO/GEOLino. The use of second person pronouns ranks highly for both data sets, which may easily be explained by it being used for the informal German address appropriate when speaking to children. This is further corroborated by the ratio of second person verb inflections being ranked as the 13th most important measure. For Tagesschau/Logo, other implicit measures of textual cohesion based on content overlap are also informative as well as the use of causal connectives. Overall 55% of the most informative 20 measures for both data set are captured by these two categories.

The other feature groups are less frequently represented, but provide some interpretable insights into the data. First, both data sets show indications of differences in the degree of nominalization used in language targeting adults and children: For GEO/GEOLino, complex noun phrases per t-unit and finite clause are highly informative as well as the use of the nominalization suffix *-ion*. On Tagesschau/Logo, genitive case, determiners per noun phrase, and the percentage of compound nouns indicate a similar relevance of differences regarding the organisation of the nominal domain. Lexical and sentential complexity seems to be less homogeneous for the distinction of adult and child targeted language across data sets: There are two measures of lexical complexity assessing word length in syllables and the semantic inter-relatedness of words ranked high for GEO/GEOLino, while on Tagesschau/Logo, lexical diversity and verb variation are particularly informative. For sentential complexity, constituency tree complexity, the average length of the middle field, and the use of prepositional phrases per t-unit are particularly informative on GEO/GEOLino. On Tagesschau/Logo, parse tree height and the use of conjunctive clauses are relevant. Cognitive measures do not seem to play an important role on either data set, except for the sum of longest dependencies per clause on Tagesschau/Logo.

Overall, these results clearly show that for both data sets the distinction between target audiences is not just made based on surface modifications such as sentence or word length. In fact, these measures do not occur among the most informative measures at all. Rather, measures of language use and cohesion are predominantly informative for the distinction of adult and child targeting texts, but also measures of phrasal, sentential, lexical, and morphological complexity. The adjustment of the data to their audience observed here thus seems to be more linguistically refined than that found in the *ReadingDemands* textbook data, where Berendes et al. (2017) found only few adjustment across dimensions.

6 Study 2: Can we successfully model readability for German, also across data sets?

6.1 Set-Up

Our second objective is the design of a robust model of educational media language that distinguishes robustly between language targeting adults and children across corpora and genres. For this, we train two binary Sequential Minimal Optimization (SMO) support vector classifier (Platt, 1998) with linear kernels using the WEKA machine learning toolkit (Hall et al., 2009). Each model is tested i) on the same corpus it is trained on, using 10-folds cross-validation, and afterwards ii) on the other data set for cross-corpus testing after training on the full data set. For model performance evaluation, we report classification accuracy and the classification confusion matrices, and random baselines as reference point.

6.2 Results and Discussion

Table 6 shows the accuracy of our SMO models on both data sets and compares them with a random baseline. Both models clearly outperform the baseline of 50.0%. On GEO/GEOLino_S, the performance is comparable to the performance observed by Hancke et al. (2012) on the original GEO/GEOLino data.¹¹

As Table 7a shows, erroneous classifications are roughly balanced across both classes, showing that the model does not prefer one class over the other. When training a model using only the 20 most informative measures identified in Study 1, we reach an accuracy of 85.1%, i.e., the additional measures only account only for 3.3%.¹² When testing the models on the Tagesschau/Logo corpus, accuracy increases to 98.8% for both models. The confusion matrix for the model using 400 measures in Table 7b seems to indicate

¹¹After observing these results, we obtained the original GEO/GEOLino data set from Hancke et al. (2012) and trained and tested a model with 10-folds cross-validation on it. When using the same data, our model outperforms their best performing

Model	Training	Testing	Features	Accuracy	SD
Baseline		GEO/GEOLino _S		50.0	
		Tagesschau/Logo		50.0	
10-folds CV	GEO/GEOLino _S	GEO/GEOLino _S	400	89.4	±0.09
			20	85.1	±0.09
	Tagesschau/Logo	Tagesschau/Logo	400	99.9	±0.04
			20	99.8	±0.03
Cross-Corpus	GEO/GEOLino _S	Tagesschau/Logo	400	98.9	
			20	98.8	
	Tagesschau/Logo	GEO/GEOLino _S	400	52.2	
			20	56.7	

Table 6: Classification performance of model on GEO/GEOLino_S and Tagesschau/Logo data

↓Obs./Prd.→	Child _{GEOLino}	Adult _{GEO}	↓Obs./Prd.→	Child _{GEOLino}	Adult _{GEO}
Child _{GEOLino}	2,222	258	Child _{Logo!}	408	7
Adult _{GEO}	267	2,213	Adult _{TS}	2	419

(a) 10-folds CV on GEO/GEOLino_S (b) Cross-corpus testing on Tagesschau/Logo

Table 7: Confusion matrices for testing models with 400 features trained on GEO/GEOLino_S.

a minor tendency towards classifying *Logo!* texts as Tagesschau texts, but due to the low number of incorrect classifications this is not conclusive.

Overall, performance of both models trained on GEO/GEOLino_S on the Tagesschau/Logo data is comparable to the performance of both models trained and tested on Tagesschau/Logo with 10-folds cross-validation, although the confusion matrix for the cross-validated Tagesschau/Logo model using 400 measures does not exhibit any tendency towards predicting one class preferred over the other, as may be seen in Table 8a.

↓Obs./Prd.→	Child _{Logo!}	Adult _{TS}	↓Obs./Prd.→	Child _{Logo!}	Adult _{TS}
Child _{Logo!}	415	0	Child _{GEOLino}	2,472	8
Adult _{TS}	1	420	Adult _{GEO}	2,362	118

(a) 10-folds CV on Tagesschau/Logo (b) Cross-corpus testing on GEO/GEOLino_S

Table 8: Confusion matrices for testing models with 400 features trained on Tagesschau/Logo

The model trained and tested on Tagesschau/Logo reaches an unexpectedly high accuracy of 99.9% for using 400 measures and 99.8% when using only the 20 most informative measures reported in Study 1. Since the performance remains high when using only 20 measures and the standard deviation across folds is very low, the results seem not to be due to over-fitting. The model learns linguistic properties of the data set that generalize across. It is important to stress here that none of our measures include n-gram language models or any other lexical content features but only complexity measures aggregated over each document.¹³

model with 91.1%, confirming that our approach is in fact competitive with the state of the art.

¹²We do not show the confusion matrices for the models with 20 features, because they are equivalent to the matrices in Table 7. The same holds for the models tested on Tagesschau/Logo and their matrices in Table 8.

¹³Content features are problematic since they can pick up recurring phrases that are characteristic of particular media outlets rather than generalizable linguistic complexity characteristics. E.g., the *Tagesschau* always starts with the greeting “Hier ist das Erste Deutsche Fernsehen mit der Tagesschau.” (*Here is the first public German TV channel with the daily news.*)

When testing the models trained on the Tagesschau/Logo data set on the GEO/GEOLino_S data, it becomes apparent that the characteristics learned from the Tagesschau/Logo data set do not generalize, with the model based on 400 measures performing only marginally above chance, and the model using the 20 measures performing slightly better with 56.2%. When considering the confusion matrix for this model in Table 8b, we see that most texts are classified as GEOLino texts, irrespective of whether they belong to GEO or GEOLino. The Tagesschau/Logo trained models do not generalize well to the other adult/child corpus. Since the model trained on GEO/GEOLino_S is highly successful when tested on Tagesschau/Logo, this cannot be due to an actual lack of generalizable differences in the linguistic characteristics of the adult and child targeting texts contained in both data sets. One possible reason for these results may be that, as Study 1 showed, the measures are considerably more informative on Tagesschau/Logo than on GEO/GEOLino_S. It could be, that the differences between the news subtitles designed for different target audiences are more extreme than those observable for the GEO magazines. This would explain the surprisingly good performance of the GEO/GEOLino_S model on the Tagesschau/Logo data, which would then be easier to distinguish, while also accounting for the poor performance in the opposite case.

7 Summary and Outlook

We presented a study of the difference between German targeting adults and children, as far as we know the most broadly based linguistic complexity analysis to date. We created and analyzed a novel data set compiled from German news subtitles that consists of news broadcasts for adults and children from the same days, ensuring a relatively parallel selection of topics. We compared this with a newly compiled GEO/GEOLino corpus consisting of online articles of two magazines for adults and children by the same publisher discussing the similar topics. Based on these two data sets, we presented within-corpus (10-fold CV) and cross-corpus experiments and built binary classification models of German educational media text readability that perform with very high accuracy across both data sets. The model is based on a broad range of features that are highly informative for both data sets. This model is a valuable contribution since i) it is based on a considerably broader data basis than previous approaches to German readability, and ii) it successfully generalizes across the data sets, illustrating surprising robustness across rather different text types. The approach presented thus extends the state-of-the-art in Hancke et al. (2012) in terms of the breadth of features integrated and the accuracy and generalizability of the model – and provides two new data sources for this line of research.

The paper also contributes some new insights into the linguistic characteristics of German media language targeting adults and children. Since all the language is produced by adults, it is not necessarily clear how well it is in fact adjusted to the target audience. As demonstrated by Berendes et al. (2017), German textbook publishers indeed do not seem to be adjusting the complexity of the language used according to school type and grade level in any systematic way. Our results for educational media language indicate, that i) both data sets are successfully and broadly adapted towards their target audiences; and ii) that they form two distinct, cross-corpus generalizable constructs of German educational media language for children and adults. In a next step, we plan to test to which extent this linguistically diverse and generalized construct matches the language competence of the intended children target group by comparing it with the Karlsruhe Children's Text corpus (Lavalley et al., 2015). We also plan to further investigate the linguistic properties of our two data sets. In particular, the Tagesschau/Logo data set requires further statistical and qualitative analyses to investigate why its linguistic characteristics generalize well across all folds of the data set itself but not across GEO/GEOLino. We also plan to conduct more analyses of the informativeness of the different complexity feature groups for the target audience distinction.

Acknowledgements

We would like to thank Peter Lindner of the NDR and the editorial office of *ARD Aktuell* for providing us with the news subtitles of the *Tagesschau* broadcasts and Christiane Müller of the ZDF for giving us access to the subtitles from *Logo!*.

References

- Stefanie Albert, Jan Anderssen, Regine Bader, Stephanie Becker, Tobias Bracht, Sabine Brants, Thorsten Brants, Vera Demberg, Stefanie Dipper, Peter Eisenberg, Silvia Hansen, Hagen Hirschmann, Juliane Janitzek, Carolin Kirstein, Robert Langner, Lukas Michelbacher, Oliver Plaehn, Cordula Preis, Marcus Pußel, Marco Rower, Bettina Schrader, Anne Schwartz, George Smith, and Hans Uszkoreit, 2003. *TIGER Annotationschema*. Universität des Saarlandes and Universität Stuttgart and Universität Potsdam.
- Karin Berendes, Sowmya Vajjala, Detmar Meurers, Doreen Bryant, Wolfgang Wagner, Maria Chinkina, and Ulrich Trautwein. 2017. Reading demands in secondary school: Does the linguistic complexity of textbooks increase with grade level and the academic orientation of the school track? *Journal of Educational Psychology*.
- Bernd Bohnet and Joakim Nivre. 2012. A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1455–1465, Jeju Island, Korea, July. Association for Computational Linguistics, Association for Computational Linguistics.
- Doreen Bryant, Karin Berendes, Detmar Meurers, and Zarah Weiß. 2017. Schulbuchtexte der Sekundarstufe auf dem linguistischen Prüfstand. Analyse der bildungs- und fachsprachlichen Komplexität in Abhängigkeit von Schultyp und Jahrgangsstufe. In Mathilde Hennig, editor, *Linguistische Komplexität – ein Phantom?* Stauffenburg Verlag, Tübingen.
- Marc Brysbaert, Matthias Buchmeier, Markus Conrad, Arthur M. Jacobs, Jens Bölte, and Andrea Böhl. 2011. The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in german. *Experimental Psychology*, 58:412–424.
- Bram Bulté and Alex Housen. 2014. Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26(0):42 – 65. Comparing perspectives on L2 writing: Multiple analyses of a common corpus.
- Jeanne S. Chall and Edgar Dale. 1995. *Readability revisited: the new Dale-Chall Readability Formula*. Brookline Books.
- Xiaobin Chen and Detmar Meurers. 2016. CTAP: A web-based tool supporting automatic complexity analysis. In *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity*, Osaka, Japan, December. The International Committee on Computational Linguistics.
- Xiaobin Chen and Detmar Meurers. 2017. Word frequency and readability: Predicting the text-level readability with a lexical-level attribute. *Journal of Research in Reading*. JRIR-2017-01-0006.R1.
- Scott A. Crossley, Jerry Greenfield, and Danielle S. McNamara, 2008. *Assessing text readability using cognitively based indices*, pages 475–493. Teachers of English to Speakers of Other Languages, Inc. 700 South Washington Street Suite 200, Alexandria, VA 22314.
- Scott A. Crossley, David B. Allen, and Danielle McNamara. 2011. Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language*, 23(1):84–101, April.
- William H. DuBay. 2004. *The Principles of Readability*. Impact Information, Costa Mesa, California.
- Duden. 2009. *Deutsche Grammatik*, volume 4. Dudenverlag, 4 edition.
- Rod Ellis and Gary Barkhuizen. 2005. *Analysing learner language*. Oxford University Press.
- B. Janghorban Esfahani, A. Faron, K. S. Roth, P. P. Grimminger, and J. C. Luers. 2016. Systematic readability analysis of medical texts on websites of german university clinics for general and abdominal surgery. *Zentralblatt für Chirurgie*, 141(6):639–644.
- Lijun Feng, Martin Jansche, Matt Huenerfauth, and Noémie Elhadad. 2010. A comparison of features for automatic readability assessment. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)*, Beijing, China.
- Thomas François and Cedrick Fairon. 2012. An “AI readability” formula for French as a foreign language. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*.
- Edward Gibson. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, pages 95–126.

- Petronella Grootens-Wiegers, Martine C. De Vries, Tessa E. Vossen, and Jos M. Van den Broek. 2015. Readability and visuals in medical research information forms for children and adolescents. *Science Communication*, 37(1):89–117.
- Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA data mining software: An update. In *The SIGKDD Explorations*, volume 11, pages 10–18.
- Julia Hancke, Detmar Meurers, and Sowmya Vajjala. 2012. Readability classification for German using lexical, syntactic, and morphological features. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, pages 1063–1080, Mumbai, India.
- Julian Heister, Kay-Michael Würzner, Johannes Bubenzer, Edmund Pohl, Thomas Hanneforth, Alexander Geyken, and Reinhold Kliegl. 2011. dlexDB - eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau*, 62:10–20.
- Alexis Housen, Folkert Kuiken, and Ineke Vedder. 2012. Complexity, accuracy and fluency: Definitions, measurement and research. In Alex Housen, Folkert Kuiken, and Ineke Vedder, editors, *Dimensions of L2 Performance and Proficiency*, Language Learning & Language Teaching, pages 1–20. John Benjamins.
- Kellogg W. Hunt. 1970. Do sentences in the second language grow like those in the first? *TESOL Quarterly*, 4(3):195–202.
- J. Peter Kincaid, Robert P. Fishburne, Richard L. Rogers, and Brad S. Chissom. 1975. Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease formula) for Navy enlisted personnel. Research Branch Report 8-75, Naval Technical Training Command, Millington, TN.
- Rémi Lavalley, Kay Berkling, and Sebastian Stüker. 2015. Preparing children’s writing database for automated processing. In *LTLT@ SLaTE*, pages 9–15.
- Xiaofei Lu. 2010. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4):474–496.
- Xiaofei Lu. 2012. The relationship of lexical richness to the quality of ESL learners’ oral narratives. *The Modern Languages Journal*, pages 190–208.
- Sarah E. Petersen and Mari Ostendorf. 2009. A machine learning approach to reading level assessment. *Computer Speech and Language*, 23:86–106.
- Slav Petrov and Dan Klein. 2007. Improved inference for unlexicalized parsing. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 404–411, Rochester, New York, April.
- John C. Platt. 1998. Sequential minimal optimization: A fast algorithm for training support vector machines. Technical Report MSR-TR-98-14, Microsoft Research.
- Anna N. Rafferty and Christopher D. Manning. 2008. Parsing three German treebanks: lexicalized and unlexicalized baselines. In *Proceedings of the Workshop on Parsing German, PaGe ’08*, pages 40–46, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Marga Reis. 2001. Bilden Modalverben im Deutschen eine syntaktische Klasse? *Modalität und Modalverben im Deutschen*.
- Robert Reynolds. 2016. Insights from Russian second language readability classification: complexity-dependent training requirements, and feature evaluation of multiple categories. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 289–300, San Diego, CA, June. Association for Computational Linguistics.
- Karsten Schmidt. 2016. Der graphematische Satz. *Zeitschrift für germanistische Linguistik*, 44(2):215–256.
- Sowmya Vajjala and Detmar Meurers. 2012. On improving the accuracy of readability classification using insights from second language acquisition. In Joel Tetreault, Jill Burstein, and Claudial Leacock, editors, *In Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 163–173, Montréal, Canada, June. Association for Computational Linguistics.
- Sowmya Vajjala and Detmar Meurers. 2014. Assessing the relative reading level of sentence pairs for text simplification. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 288–297, Gothenburg, Sweden, April. ACL, Association for Computational Linguistics.

- Heike vom Orde. 2015. Kindernachrichten im Fernsehen. Eine Zusammenfassung zentraler Forschungsergebnisse zum Format logo! *Television*, 28/2015/2:40–42.
- Tim Vor der Brück, Sven Hartrumpf, and Hermann Helbig. 2008. A readability checker with supervised learning using deep syntactic and semantic indicators. *Informatica*, 32(4):429–435.
- Karl Woodmansey. 2010. Readability of educational materials for endodontic patients. *Journal of Endodontics*, 36:1703–1706.
- Ramon Ziai and Detmar Meurers. 2018. Automatic focus annotation: Bringing formal pragmatics alive in analyzing the Information Structure of authentic data. In *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, New Orleans, LA. ACL. To appear.

Appendix A. Definitions of Linguistic Units

Clauses are all maximal projections of finite verbs and elliptical constructions with sentential status (i.e. all sub-trees tagged with *S*), as well as *to* infinitives that have a sentential status (*satzwertige zu Infinitive*).

Complex t-units are t-units that include subordinate clauses.

Conjunctive clauses are all dependent clauses that are introduced by a subordinating conjunction such as *dass*, *weil*, or *wenn*.

Dependent clauses with conjunction are all conjunctive clauses, but also interrogative and relative clauses. Dependent clauses without conjunction are mostly dependent main clauses, such as *Ich weiß*, *es ist spät*.

(Graphematic) sentences are strings of at least one token that are ended by sentence ending punctuation marks: *!*, *.*, *?*. There is a broad discussion on alternative sentence definitions, see for example Schmidt (2016) for a more elaborate theoretical account. However, since sentences are identified by sentence segmentation tools, which are primarily based on punctuation, sentences are always defined as graphematic sentences.

Half modals are *haben*, *sein*, *scheinen*, *drohen*, *versprechen*, if they govern an infinitive with *zu* (Duden, 2009, §101), e.g. *ist zu machen*, *droht zu schneien*.

Lexical words are all nouns, adjectives, adverbs, foreign words, numbers, main verbs, and modal verbs. Note that there is an ongoing discussion on whether modals actually qualify as lexical words (Reis, 2001), hence there is also a subset of lexical words excluding modals employed throughout the system.

Parts-of-Speech are operationalized following the Tiger POS tags (Albert et al., 2003, 121).

Quasi passives are *bekommen*, *erhalten* or *kriegen* if they govern a past participle (Duden, 2009, §179), e.g. *bekommt gemacht*, *kriegt eröffnet*.

T-units are “one main clause plus any subordinate clause or non clausal structure that is attached to or embedded in it” (Hunt, 1970, 4).

Appendix B. Example Extracts from *Tagesschau* and *Logo!* subtitles

Report on New Years shooting in Istanbul by *Tagesschau*, extracted from the subtitles for the broadcast on 01.01.2017, 20:00.

In der Türkei ist der Jahreswechsel von einem Anschlag in Istanbul überschattet: Mindestens ein bewaffneter Angreifer drang in einen Nachtclub ein und schoss um sich. 39 Menschen wurden getötet und mehr als 60 verletzt. Unter den Todesopfern sind zahlreiche Ausländer. Ob Deutsche betroffen sind, ist unklar. Die Suche nach dem Täter dauert an, bekannt hat sich niemand. Das Attentat ereignete sich im europäischen Teil Istanbuls. Dort liegt direkt am Bosphorus der Club "Reina", der bei Prominenten beliebt ist. Nur eine Stunde währte in der Türkei die Hoffnung, 2017 könnte ruhiger werden als 2016, das von Bombenanschlägen geprägt war. Doch um 1.15 Uhr Ortszeit macht im Istanbuler Nachtclub "Reina" ein Attentäter mit einem Gewehr Jagd auf Gäste einer Silvesterparty. Zuvor wurde vorm Club ein Polizist erschossen. Der Täter konnte fliehen, eine Großfahndung läuft. Bis zu 800 Personen sollen sich in der Diskothek aufgehalten haben. Gäste berichten, Panik sei ausgebrochen. Einige Besucher sollen in den Bosphorus gesprungen sein. Unter den Toten und Verletzten sind Ausländer. Bekannt hat sich niemand zu der Tat. Türkische Medien vermuten den IS hinter dem Terrorakt. Die Regierung verhängte eine Nachrichtensperre. Wir lassen uns vom Terror nicht beirren. Was hier passierte, kann morgen an einem anderem Ort geschehen. Es gibt keine Garantien. Der Nachtclub "Reina" liegt am Bosphorus, im Stadtteil Ortaköy. Er ist der berühmteste der Türkei, teuer und bei Touristen beliebt. Die Sicherheitsvorkehrungen waren landesweit erhöht worden. In Istanbul waren 17.000 Polizisten im Einsatz. Trotz Großaufgebot der Polizei, hochaktiver Geheimdienste, Ausnahmezustand und markiger Politikerworte: Die Sicherheitslage in der Türkei spitzt sich zu. Beängstigende Aussichten für Wirtschaft und Menschen.

Report on New Years shooting in Istanbul by *Logo!*, extracted from the subtitles for the broadcast on 02.01.2017, 19:50 (no broadcast on 01.01.2017).

In der türkischen Großstadt Istanbul hat es an Silvester einen Anschlag in einer Disco gegeben. Ein Mann stürmte mit einem Gewehr in den Club und hat 39 Menschen getötet, darunter auch zwei Männer, die in Deutschland gelebt haben. Die türkische Polizei sucht jetzt nach dem Täter. Er ist seit dem Anschlag auf der Flucht. Auch am zweiten Tag nach dem Anschlag kamen viele Menschen an die Polizeiabspernung, um Blumen für die Opfer niederzulegen. Der Terrorist stürmte dort in der Silvesternacht mit einem Gewehr in die Disco. Ich war völlig geschockt, konnte mich nicht bewegen. Der Täter schoss erst auf einen Polizisten und dann auf die Gäste. Wir hörten plötzlich Schüsse, da sind wir raus aus dem Ballsaal auf die Terrasse und haben uns dort versteckt. Im Internet behauptet die Terrorgruppe IS, Islamischer Staat, dass sie hinter dem Anschlag stecke. Die Kämpfer dieser Terrorgruppe wollen, dass alle Menschen nach ihren strengen religiösen Regeln leben. Wer sich nicht daran hält, wird sogar umgebracht. Besonders aktiv ist der IS in Teilen von Syrien und dem Irak. Beide Länder grenzen an die Türkei. Dort haben die Kämpfer in letzter Zeit schon öfter Anschläge verübt. In der ganzen Türkei sucht die Polizei jetzt nach dem Attentäter. Acht Verdächtige wurden schon festgenommen. Auf logo.de könnt ihr mehr zur Terrorgruppe Islamischer Staat lesen und da gibt es auch viele Infos zu unserem nächsten Thema.

Appendix C. Example Articles from GEO and GEOLino

GEO article titled “Was ist ein Planet?” (What is a Planet?).¹⁴ It discusses criteria celestial bodies need to fulfill to be considered a planet.

Lange bezeichneten Menschen alle Lichtpunkte, die über den Nachthimmel wanderten, als Planeten (griech. *planáomai* = umherirren) – gleich, ob es sich um Venus, Mars, Mond oder Asteroiden handelte. In der Neuzeit durften den Titel nur noch die großen Himmelskörper tragen, die um die Sonne kreisten, aber keine Monde waren – also nicht ihrerseits einen anderen Planeten umrundeten. Als Astronomen von 1992 an in den Randbezirken des Sonnensystems immer neue Objekte entdeckten, manche ähnlich groß wie Pluto (bis dahin der neunte Planet), sah sich die Internationale Astronomische Union genötigt, erstmals zu definieren, was ein Planet genau ist. Nach heftigen Diskussionen beschlossen die Astronomen 2006 die Resolution B5. Demnach muss ein Planet drei Kriterien erfüllen: Er muss um die Sonne kreisen. Er muss ausreichend Masse aufweisen, sodass er unter seiner eigenen Schwerkraft eine nahezu runde Form angenommen hat. Und er muss die Umgebung seiner Umlaufbahn freigeräumt haben. Objekte, die ihm auf seiner Bahn nahekommen, “schluckt” er in einer Kollision oder schleudert sie in einen anderen Orbit. Pluto, Eris und andere große Himmelskörper zählen nun zu den Zwergplaneten, da sie es nicht schaffen, ihre Bahn zu bereinigen, sondern sie sich mit anderen Objekten teilen. Damit kreisen nach derzeitigem Stand acht Planeten um die Sonne. Die Astronomen unterteilen sie in die vier terrestrischen Planeten Merkur, Venus, Erde, Mars (sie werden wegen ihrer festen Oberfläche häufig steinige Planeten genannt) und in die vier jovianischen – jupiterähnlichen – Planeten Jupiter, Saturn, Uranus, Neptun (aufgrund ihrer Zusammensetzung oft als Gasplaneten oder Gasriesen bezeichnet). Wobei Uranus und Neptun manchmal auch als “Eisriesen” beschrieben werden, da sie weniger Wasserstoff als Jupiter und Saturn enthalten, dafür mehr gefrorenes Methan, Wasser und Ammoniak.

GEOLino article titled “Sieben erdähnliche Planeten entdeckt” (Seven Earth-Like Planets Discovered).¹⁵ It reports on the discovery of seven new planets that orbit Trappist-1.

Dass neue Planeten entdeckt werden, ist erstmal nichts ungewöhnliches. Doch der Fund dieser sieben sogenannten Exoplaneten (Planeten wie Kepler-452b, die sich um einen Stern - außerhalb des Einflusses unserer Sonne - bewegen) ist etwas ganz Besonderes: Denn sechs der neu entdeckten Planeten liegen in einer Temperaturzone, in der Leben möglich ist. Auf den meisten Planeten ist es entweder kochend heiß oder eiskalt - schwierige Bedingungen für die Entwicklung von Leben. Die Sonne der Exoplaneten, der Zwergstern Trappist-1, ist viel kleiner als die Sonne unseres Sonnensystems: Trappist-1 besitzt nur acht Prozent der Masse unserer Sonne und zwölf Prozent ihres Durchmessers. Auf drei der entdeckten Exoplaneten könnte sogar Wasser existieren, denn ihr Abstand zur Zwergsonne liegt in einem Temperaturbereich, in dem Wasser weder gefrieren noch verdampfen würde. Hier wäre also eine Art von Leben möglich, wie wir es auf unserer Erde kennen.

Die sieben Planeten haben in etwa die Größe unserer Erde und sind wahrscheinlich Gesteinsplaneten. Sie alle umkreisen ihre Sonne, den Stern Trappist-1, der knappe 40 Lichtjahre (ein Lichtjahr ist die Strecke, die Licht in einem Jahr zurücklegt) von uns entfernt im Sternbild Wassermann liegt. Weil die Sonne des Trappist-1-Systems so klein ist, können die Planeten diese wesentlich schneller umkreisen als wie es in unserem Sonnensystem möglich ist. Die sechs Planeten, die dem Zwergstern am nächsten sind, umrunden ihn in eineinhalb bis zwölf Tagen. Sie haben damit einen engeren Orbit als der Merkur um die Sonne.

¹⁴<http://www.geo.de/wissen/weltall/15396-rtkl-definitionssache-was-ist-ein-planet>, accessed 11.06.18, 16:06.

¹⁵<http://www.geo.de/geolino/wissen/weltraum/sieben-erdaehnliche-planeten-entdeckt>, last accessed 11.06.18, 16:06.