Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 7123–7131
Marseille, 11–16 May 2020

CTAP for Italian: Integrating Components for the Analysis of Italian into a Multilingual Linguistic Complexity Analysis Tool

Nadezda Okinina, Jennifer-Carmen Frey, Zarah Weiss
Institute for Applied Linguistics, Eurac Research, Viale Druso 1, 39100 Bolzano, Italy
Department of Linguistics, University of Tübingen, Wilhelmstrasse 19, 72074 Tübingen, Germany
{Nadezda.Okinina, JenniferCarmen.Frey}@eurac.edu, zweiss@sfs.uni-tuebingen.de

Abstract

Linguistic complexity research being a very actively developing field, an increasing number of text analysis tools are created that use natural language processing techniques for the automatic extraction of quantifiable measures of linguistic complexity. While most tools are designed to analyse only one language, the CTAP open source linguistic complexity measurement tool is capable of processing multiple languages, making cross-lingual comparisons possible. Although it was originally developed for English, the architecture has been extended to support multi-lingual analyses. Here we present the Italian component of CTAP, describe its implementation and compare it to the existing linguistic complexity tools for Italian. Offering general text length statistics and features for lexical, syntactic, and morpho-syntactic complexity (including measures of lexical frequency, lexical diversity, lexical and syntactical variation, part-of-speech density), CTAP is currently the most comprehensive linguistic complexity measurement tool for Italian and the only one allowing the comparison of Italian texts to multiple other languages within one tool.

Keywords: linguistic complexity, readability, Italian, text analysis, cross-lingual analysis

1. Introduction

Linguistic complexity is a core construct in Second Language Acquisition (SLA) research, where Complexity, Accuracy, and Fluency, also known as the CAF triad, are often used to characterize language performance (Housen and Kuiken, 2009). Over the last decade, a broad variety of complexity measures has been proposed to characterize language proficiency and its development, text readability, and writing quality (Vajjala and Meurers, 2012; Bulté and Housen, 2014; Crossley and McNamara, 2014; De Clercq and Housen, 2019). Many of these linguistic complexity measures are applicable across various languages. At the same time, advances in Natural Language Processing (NLP) allow to automatically extract the measures for a variety of languages. In the face of these advances, a series of automatic complexity analysis approaches have been presented. There are approaches for English (McNamara and Graesser, 2012; Chen and Meurers, 2016), German (Weiss and Meurers, 2019a), French (Francois and Miltsakaki, 2012), Swedish (Pilan et al., 2016) and Portuguese (Aluisio et al., 2010) containing similar - yet not completely overlapping - sets of complexity measures.

With the creation of quantifiable operationalizations for aspects of linguistic complexity (e.g. lexical diversity) that can be calculated automatically for various languages, cross-lingual analyses can be envisioned. However, analyses using the same measures for texts in various languages are still rare. While the comparability of the measures used needs to be ensured from a theoretical perspective, often the extraction of linguistic complexity indices for various languages is also technically limited. Although various language-dependent tools exist that extract linguistic complexity indices from text, they usually provide very different feature sets. Besides, they are often based on different assumptions and use different technologies that ultimately lead to different values even for the same features. Using one single tool to extract those features would, however, substantially simplify analysis workflows.

For this reason, we extended the Common Text Analysis Platform (CTAP) (Chen and Meurers, 2016) to support the analysis of Italian. We transferred the linguistic complexity measures already provided for English and German by integrating an Italian text processing pipeline and added further features frequently used in the Italian context (e.g. the Gulpease readability measure). With these developments, the platform now supports three European languages (Italian, German, and English), with a total of 154 linguistic complexity features that can be extracted for all three of them. The Italian component of CTAP includes with 253 features more complexity measures than the other existing tools for Italian, providing a flexible feature extraction that is not limited to specific research questions.

In this article, we describe the Italian component of CTAP. After a brief review of related complexity research in Section 2, we situate CTAP among the already existing tools for Italian linguistic complexity measurement, introducing the main aims and research focuses of the tools (Section 3). Subsequently, we present the general architecture of CTAP, explain some implementation details of its Italian component, list the linguistic complexity measures it offers and describe the quality control mechanisms we used (Section 4). Next, we compare the characteristics of CTAP with the characteristics of other linguistic complexity measurement tools for Italian (Section 5), before we conclude our article also pointing out future work (Section 6).

2. Linguistic Complexity Research

As a central dimension of (second) language performance, complexity has been extensively researched in the context of assessing second language proficiency and development (Crossley and McNamara, 2014; Kyle, 2016; Bulté and Housen, 2014). However, complexity measures have also been shown to be beneficial for other tasks such as readability assessment (Vajjala and Meurers, 2012; Feng et al., 2010; Chen and Meurers, 2018), first language academic writing acquisition (Crossley et al., 2011; Weiss and Meurers, 2019b), and the evaluation of teachers grading behaviour (Vögelin et al., 2019; Weiss et al., 2019). Most of this work has focused on English, but especially in recent years, the scope of complexity research has been broadened towards other languages such as German (Weiss and Meurers, 2019a), Swedish (Pilan et al., 2016), Russian

(Reynolds, 2016), French (Francois and Miltsakaki, 2012), and Italian (Brezina and Pallotti, 2019). With this broadening across languages, the types of complexity measures that are being investigated have also been extended, thus overcoming the so far often reductionist approach to complexity (cf. Housen et al., 2019), which focuses nearly exclusively on lexical and syntactic complexity. Overcoming this reductionist approach has become one of the central goals in complexity research (Housen et al., 2019; Paquot, 2019). New complexity measures are being proposed and tested for various languages. For example, new measures of morphological complexity have been used to characterize the developmental trajectories of second language (L2) spoken French and English (De Clercq and Housen, 2019) and to distinguish between native and L2 speech for Italian and English (Brezina and Pallotti, 2019). Both studies find their measures of morphological complexity to be highly informative for the respective non-English languages.

However, these advances in broadening the scope of complexity research are not necessarily accompanied by efforts to make the newly proposed measures accessible to a broader audience. Researchers trying to navigate through the increasing collection of complexity measures find themselves often at a loss. Few tools provide comprehensive collections of complexity measures and these are typically language dependent rather than facilitating complexity analyses across languages.

3. Linguistic Complexity Measurement Tools for Italian

In the context of Italian language research, a number of feature extraction tools has been developed over the last years, differing in their intended research aims and domains. To our knowledge, there are three tools for measuring linguistic complexity of Italian texts apart from CTAP introduced in this article: READ-IT (Dell'Orletta et al., 2011), Coease (Tonelli et al., 2012), and Tint 2.0 (Aprosio and Moretti, 2018). All three tools originated in computational linguistic research on text readability and text simplification. Early readability research has focused on a small set of superficial text characteristics such as word and sentence length which could be employed in simple readability formulae, see DuBay (2004) for an overview. However, the use of linguistically more informed complexity measures has been shown to be more appropriate to model readability (Vajjala and Meurers, 2012; Feng et al., 2010; Chen and Meurers, 2018) and since then, complexity measures have become an important component in readability assessment research.

READ-IT was designed to study text simplification approaches for readers with low literacy skills or mild cognitive impairment and has been used, for example, to analyse the readability of informed consent forms in the public health sector (Venturi et al., 2015), or to explore linguistic features of Italian fictional prose across textual genres and readability levels (Dell'Orletta et al., 2013). It focuses on classical readability measures for Italian, including some other surface features such as, for example, text, sentence or token length.

Coease was created to analyse text complexity and readability in educational settings and focuses on the evaluation of textual difficulty according to different educational levels. It was built along the model of Coh-Metrix (McNamara and Graesser, 2012), a popular linguistic complexity measurement tool for English often used in second language acquisition and writing research.

By incorporating complexity measures into the all-inclusive NLP suite for Italian, Tint 2.0, its authors offer open source implementations of some of the complexity measures present in Coease and READ-IT. It thus mirrors the research foci of the two other tools.

By adapting CTAP to Italian, we aimed to provide a complexity feature extraction tool with a broader and more generic set of features than the three existing tools for Italian, without focusing on aggregate readability indices. Both READ-IT and Coease make extensive use of such measures and besides offer an interpretation of the measures obtained for a text in terms of how they compare with a representative sample of a specific text type. CTAP does not use such reference corpora for giving interpretations but allows researchers to use their own corpora for comparison. This keeps the tool as flexible as possible serving a wide range of research purposes. With its flexible and easily extendible architecture, CTAP furthermore allows to exchange individual parts of the processing pipeline, reconfigure settings and parameters and integrate new features if needed. Finally, the tool allows to extract the same measures for different languages, making it interesting for cross-lingual analysis.

4. CTAP and Its Extension to Italian

The Common Text Analysis Platform CTAP (Chen and Meurers, 2016) is a web-based quantitative linguistic feature extraction tool for measures of linguistic complexity. Contrary to other tools that provide pre-defined analysis set-ups for individual texts, CTAP is fully configurable. It is not limited to any specific task but can be used in any project that requires the extraction of quantitative linguistic features out of written texts.

4.1 General Architecture of CTAP

CTAP is based on the Unstructured Information Management (UIMA) framework (Ferrucci et al., 2004) that facilitates the addition of new components to the already existing software architecture. The analysis pipeline for the complexity measures is separated into two components:

(1) Annotators of basic linguistic structures such as letters, syllables, tokens, lemmas, POS categories, sentences, and syntactic structures. These take plain text or the output of other annotators as input and generate annotations.

(2) Analysis engines that generate complexity features' values. These take as input the annotations produced by the annotators of linguistic structures and generate the values for individual complexity measures.

This division of the analysis architecture makes the integration of new languages as well as new complexity measures very feasible. As the complexity analysis engines use the output of linguistic annotators, by adding linguistic analysers for a new language, a wide range of complexity measures can be obtained without further modifications, except for inserting the language code into the corresponding feature descriptors. On the other hand, when a new complexity measure analysis engine is implemented, it can be applied to all the languages that already have their linguistic annotators included into the platform. However, when the output of a linguistic annotator is language-specific (as that of a POS tagger or a syntactic parser, for example), new parameters need to be provided to the complexity analysis engines that use their values through XML feature descriptors. Furthermore, certain complexity measures depend on language specific external resources such as word lists or reference corpora (e.g. lexical sophistication features), which also have to be integrated into CTAP for every new language.

Originally developed for analysing English by Chen and Meurers (2016), the platform was later extended to support multilingual analysis by Zarah Weiss who also integrated a series of German complexity features into CTAP, which have been successfully used for broad linguistic modelling of German in a variety of contexts (Weiss and Meurers, 2018, 2019a,b). Our contribution consists in integrating the linguistic annotators for Italian into the tool and in adapting the existing feature sets to Italian. We also implemented several new analysis engines including:

- MTLD and HD-D, two commonly used measures of linguistic diversity (Jarvis, 2007) that can be used for all three languages,
- The Flesch-Kincaid grade level (Kincaid et al., 1975) and the Gulpease index (Lucisano and Piemontese, 1988) as readability measures for English and Italian respectively.

4.1.1 NLP Components Integrated into CTAP for the Analysis of the Italian Text

Like in the English and German components of CTAP, we use Open NLP for sentence splitting. For tokenisation and lemmatisation, we use Tint 0.2, a Maven distribution of the all-inclusive NLP suite for Italian in its first version (Aprosio and Moretti, 2016). As for the part-of-speech (POS) tagging, we use the Open NLP POS Maxent tagger[1] that reports 97.56% of accuracy. For the syntactic analysis, we use Tint 0.2 that produces Universal Dependency trees and is reported to give 84.67 LAS (labelled attachment score) and 87.05 UAS (unlabelled attachment score)[2]. As there was no syllable annotation available in the latest version of Tint referenced in the Maven repository, we wrote our syllable annotator transcribing and extending the code of the Perl module Lingua::IT:Hyphenate by Aldo Calpini[3].

4.2 Complexity Measures for Italian Available in CTAP

In its current state, the Italian component of CTAP contains 253 indices of linguistic complexity, 154 of which are also available for English and German[4].

The implemented measures are distributed among the following groups, four in total:

4.2.1 Lexical Features

There are various types of lexical features in CTAP:

- Number and percentage of tokens and word types with two or more syllables (4 features)
- Mean token length and its standard deviation in letters and syllables (4 features)
- Lexical sophistication (74 features)
- Lexical diversity (or richness) (9 features)
- Lexical variation (9 features)

The lexical sophistication features are calculated separately for all words, lexical words and function words, and each of them is based on both the SUBTLEX-IT (Crepaldi et al., 2015) and the Google Books 2012 (Lin et al., 2012) reference corpora. Lexical sophistication features include:

- 36 word frequency features: normal, logarithmic and logarithmic per million words
- 12 informativeness per million words features
- 12 familiarity per million words features
- 6 logarithmic contextual diversity features

In addition, six lexical sophistication features are based on imageability, concreteness and age of acquisition values provided by Burani et al. (2001) for 626 Italian nouns: each of these three values is calculated both for all lemmas and for unique lemmas of the text.

We also implemented a wide-spread measure of lexical sophistication for Italian which consists in calculating the proportion of words of a text that are listed in the De Mauro dictionary of basic Italian (De Mauro, 2016).

The lexical diversity (or richness) features include:

- 5 types of type-token ratio (TTR): normal TTR, root TTR, log TTR, corrected TTR, Uber TTR
- 2 types of MTLD: for tokens and lemmas
- 2 types of HD-D: for tokens and lemmas

The lexical variation features calculate the ratio of the number of different word types of a certain morpho-syntactic category to the number of all lexical tokens: nouns, verbs, adjectives, adverbs, modifiers, all lexical word types together. Verbs receive special attention and benefit from more different formulae of lexical variation, also proportionally to the number of verbs and not only to the number of all lexical tokens.

4.2.2 Syntactic Features

Syntactic features implemented in CTAP include the mean sentence length and its standard deviation in letters, syllables and tokens (6 features) and the number of syntactic constituents (40 features). The features regarding the number of syntactic constituents calculate the total number of specific syntactic constituents of a text, for example, the number of dependent clauses or conjunctions. 10 features give numbers relative to the number of sentences: the number of dependent clauses, coordinations, adjectival clause modifiers, adjectival modifiers, adverbial clauses, adverbial modifiers, appositional modifiers, attributives, auxiliaries, auxiliary passives per sentence. We plan to add more features of this type.

4.2.3 Morpho-Syntactic Features

Morpho-syntactic features implemented in CTAP are POS density features that calculate the ratio between the number of tokens belonging to certain morpho-syntactic categories to the total number of tokens, for example, the ratio of adjectives in a text.

4.2.4 Text Length Features

Basic text statistics implemented are the number of letters, syllables, tokens, word types, lemmas and sentences in the text (6 features).

[1] https://github.com/aciapetti/opennlp-italian-models/

[2] http://tint.fbk.eu/parsing.html

[3] https://metacpan.org/pod/Lingua::IT::Hyphenate

[4] The difference mainly results from differing POS sets and syntactic parsing outputs, as well as from the availability of word lists.

4.2.5 Traditional Readability Indices

The Gulpease readability index has been implemented for the Italian component of CTAP as an instance of a traditional readability index. Traditional readability indices aim to give a numerical indication of how difficult it is for an intended target group of readers to understand a given text. Gulpease (Lucisano and Piemontese, 1988) is a readability index similar to, e.g., the Flesh index (Flesch, 1948) but calibrated to model the difficulty of Italian texts for Italian native speakers at different educational levels. Contrary to other indices for Italian that are mostly adaptations of the Flesch index (e.g. the Flesch-Vacca index (Franchina and Vacca, 1986), the Gulpease index is calculated on character basis instead of syllables, to make automatic extraction of the index easier and more reliable.

4.3 Quality Control

In order to ensure the quality of the code, we implemented unit tests, comparing freshly obtained values against pre-calculated values for a sample text. This allowed us to manually verify the performance of CTAP and to guarantee the non-degradation of code during future modifications. However, as there is to date no gold standard or evaluation methodology for complexity measures, we relied on benchmark evaluations of the underlying NLP tools.

5. Comparing Linguistic Complexity Analysis Tools for Italian

In the following we describe the main differences between the available linguistic complexity measurement tools for Italian: READ-IT, Coease, Tint 2.0, and CTAP. We compare the tools along the following dimensions: first, we present the scope of the measures implemented in different tools, secondly, we give information about their source code availability and usage, next, we discuss the tools' extendibility and the difference in their units of analysis, as well as the transparency of the intermediate analysis steps. Table 1 provides a summary of our comparison.

Aspect	CTAP	Coease	READ-IT	Tint 2.0
No. of measures	253	46	32	21
Source code	open source	proprietar y	proprietar y	open source
Extendibil ity by external collaborat ors	fully extendib le	not extendible	not extendible	not for other lang-uages
Unit of analysis	corpus	text	text	corpus
Transpare ncy	no	no	yes	yes
GUI	yes	yes	yes	no

Table 1: Comparison of CTAP, Coease, READ-IT, and Tint 2.0.

5.1 Scope of the Implemented Measures

Because of their different underlying research aims, which were pointed out in Section 3, the set of implemented features differs substantially from one tool to the other. Whereas READ-IT and Coease focus strongly on readability and text simplification, CTAP, not being tailored to any specific research goal, is more generic and comprehensive than the other two.

Tint 2.0 offers the smallest number of complexity measures (21 in total), followed by READ-IT with 32 and Coease with 46 measures. CTAP is with 253 complexity measures for Italian the most comprehensive of the three tools. Since the features included in Tint 2.0 are a subset of the features included in READ-IT and Coease, we will not discuss Tint 2.0 individually in the remainder of the comparison. Only five complexity measures are present in all three tools: those are simple textual statistics, percentage of lemmas belonging to the basic vocabulary, and the Gulpease readability formula. 13 measures are offered by two tools out of three (highlighted in bold). The vast majority of measures are, however, only present in one tool.

Below we give an overview of the biggest differences in the implemented features. Table 2 gives a more detailed overview of supported features in the three investigated tools.

5.1.1 Basic counts

CTAP offers more fine-grained basic counts than the other two tools. Coease is the only one supporting paragraph counts.

5.1.2 Lexical complexity

Whereas in Coease and especially in READ-IT lexical complexity measures largely serve the purpose of defining to what extent a text may be understood by a less prepared reader, CTAP offers a wider range of generic measures for lexical complexity. The main differences between the tools are:

- For measuring lexical sophistication, Coease and READ-IT both employ the De Mauro basic Italian vocabulary[5], while CTAP apart from De Mauro also uses SUBTLEX-IT and Google Books 2012. For the measurement of familiarity, Coease uses the Italian Wikipedia as a reference corpus.
- So far, no measures of lexical abstractness have been implemented in CTAP for Italian, while this typology is represented in Coease by the mean hypernymy levels of nouns and verbs.
- CTAP is the only tool offering lexical variation per part of speech measures.
- CTAP extends the scope of lexical diversity measurement by providing different formulae for the TTR index, as well as by implementing the HD-D and MTLD measures[6] that are considered less text length dependent than the TTR (McCarthy et al., 2010).
- Unlike READ-IT, CTAP and Coease offer no overall lexical readability index, see also Section 5.2 for details on this.

[5] http://bit.ly/nuovo-demauro

[6] HD-D and MTLD measures were implemented by translating into Java their Python code by John Frens available at https://github.com/jfrens/lexical_diversity

5.1.3 Morpho-syntactic complexity

CTAP offers more fine-grained morpho-syntactic complexity indices than READ-IT (Coease providing only one) thus allowing more in-depth morpho-syntactic analysis of corpora.

5.1.4 Syntactic complexity

In terms of syntactic complexity measurement, READ-IT focuses on an in-depth analysis of subordination, Coease provides numerous indices for cohesion, causality and syntactic similarity, and CTAP specialises in calculating the number of different types of syntactic constituents and connectives.

5.1.5 Readability indices and overall textual complexity

While READ-IT and Coease both offer various readability indices and aggregated measures for overall textual complexity (e.g. lexical, syntactic, global, and base difficulty of the test), CTAP does not aim to provide such aggregate evaluation scores. Being offered a wide range of very fine-grained complexity measures, the users of CTAP have to draw their own conclusion as for the general complexity of a text. For that reason, only the popular Gulpease readability index for Italian has been implemented in CTAP.

5.2 Interpretation of Results

Both READ-IT and Coease provide task-specific interpretation utilities in their graphical user interfaces. READ-IT tells the user whether the feature values obtained for the analysed text are significantly higher or lower than for texts from a general newspaper corpus or a corpus of simplified texts. Coease does a similar comparison using texts of different educational levels as reference corpora. CTAP purposefully reports only numerical feature values, leaving the choice of selecting a reference corpus for comparison to the users, while offering them the possibility to compare not only single texts but also values obtained for whole corpora of their choice.

In addition to the feature values for each complexity measure, READ-IT presents results in form of aggregated scores judging the lexical, syntactic, global, and base difficulty. It uses readability models trained to distinguish between texts from the reference corpora with different feature sets. With the Gulpease readability index, it also provides another global measure of text readability that was obtained using reference texts.

CTAP does not provide a global readability estimate based on reference corpora or a similar interpretation of results with regard to external reference data. The tool exclusively calculates individual complexity measures and leaves it to the user to put these in an interpretative context. This is motivated by the fact that the interpretation of complexity measures can be heavily influenced by task effects. This accounts for the fact that complexity is a multi-faceted construct whose interpretation is highly context dependent. In particular, language production tasks have been shown to heavily influence complexity (e.g. Vajjala, 2018; Alexopoulou et al., 2017; Yoon, 2017), making single aggregate scores of complexity notoriously unreliable in general purpose contexts. However, we decided to include the Gulpease readability index as an additional measure in the set of Italian complexity features which may be used to gauge the overall complexity of the texts, if users have reason to assume that this measure is a good approximation of the global readability of the texts they analyze.

5.3 Source Code Availability and Usage

Among the existing linguistic complexity measurement tools for Italian, only Tint 2.0 and CTAP are open source.
READ-IT and Coease are proprietary tools. However, they provide a browser-based online demo version with a graphical user interface. Tint 2.0 on the other hand provides an open source NLP pipeline for Italian, usable via the command line or as Java library, that also offers a restricted set of complexity measures borrowed from READ-IT and Coease. The Italian component of CTAP is available open source at https://github.com/commul/ctap under the BSD license. Additionally, we maintain an online version of CTAP for free public use[7].

5.4 Extendibility

With regards to the extendibility of the tools, only CTAP is fully extendible. The proprietary tools READ-IT and Coease are not designed to be extendible in terms of features or other languages. Tint 2.0 is open source, but it is not specifically designed for being extended by external collaborators. Furthermore, the extension to other languages is not foreseen, given the tools specialization on Italian. The architecture of CTAP on the other hand allows to extend the tool to further languages and makes it possible to easily integrate new features for one or various of the supported languages.

5.5 Unit of Analysis

Apart from the complexity measures themselves, the tools differ in their flexibility regarding the unit of analysis. The graphical user interfaces for the available online demo versions of Coease and READ-IT only allow the analysis of one text at a time. While Tint 2.0 can be programmed to process various strings, CTAP was intentionally designed to analyse (sub)corpora consisting of multiple texts. Thus, its graphical interface allows to download comparative result spreadsheets and to display diagrams visualising the complexity measurements' values for different texts or corpora.

5.6 Transparency of Results

While the user interface of READ-IT visualizes intermediate results such as tokenisation, sentence splitting, POS-tagging and syntactic parsing, the other tools follow black box approaches, often only returning results as a single number. However, the possibility to check the correctness of intermediate steps and to understand the source of feature values would be crucial for researchers' trust in such feature extraction tools as well as for the interpretation of results.

[7] The tool is available for online use at https://kommul.eurac.edu/ctapWebApp/

6. Conclusion and Future Work

Linguistic complexity research being a very actively developing field, existing measures are constantly re-evaluated and new measures are proposed. It is important to be able to make those efforts available to the scientific community in a unified way. This not only helps to address current challenges in complexity research such as the overly reductionist focus on syntactic and lexical complexity measures criticized by Housen et al. (2019), but also supports researchers who do not have the technical background to implement a comprehensive set of complexity measures themselves. Furthermore, it increases the comparability and transparency of research findings.

In this paper, we have presented the Italian component of CTAP which supports the broad linguistic analysis of Italian in terms of 253 complexity measures with a subset of 154 measures being available for Italian, English and German. We have described its technical characteristics and functionalities and compared it to the other publicly available linguistic complexity measurement tools for Italian. CTAP allows for easy integration of new linguistic complexity measures and configuration of the already existing ones. The UIMA framework allows the addition of an unlimited number of complexity features to the tool if those are needed by the researcher. Collaboration is facilitated by the tool being open source and available on GitHub.

With this article we hope to provoke interest leading to collaboration and contribution to the development of new complexity measures for the languages already implemented in CTAP and of course for new languages. In the future, we would like to add new complexity measures for Italian and modify the graphical user interface in order to allow for the visualisation of intermediate analysis results. Additionally, CTAP is currently being extended to support more languages such as Dutch, Spanish and French in order to widen the scope of cross-lingual complexity research.

7. Acknowledgements

We would like to express our profound gratitude to Xiaobin Chen and Detmar Meurers of the University of Tübingen for creating the CTAP tool and for supporting us throughout the process of its adaptation to Italian. Our special thanks also go to our colleague Lorenzo Zanasi who kindly helped us in compiling lists of Italian connectives.

Linguistic complexity measure	CTAP	Co-ease	READ-IT
Basic counts			
Number of sentences	+	+	+
Number of tokens	+	+	+
Mean number of tokens per sentence	+	+	+
Mean number of paragraphs	-	+	-
Mean number of letters per token	+	-	+
Mean number of syllables per content word	-	+	-
Mean number of syllables per token	+	-	-
Mean number of sentences per paragraph	-	+	-
Standard deviation number of syllables per token	+	-	-
Standard deviation number of letters per token	+	-	-
Standard deviation number of tokens per sentence	+	-	-
Standard deviation number of letters per sentence	+	-	-
Mean number of syllables per sentence	+	-	-
Lexical complexity indices			
Number and percentage of tokens and word types with two or more syllables	+	-	-
TTR	+	+	-
TTR on the first 100 words	-	-	+
Uber, log, root, corrected TTR	+	-	-
MTLD	+	-	-
HD-D	+	-	-
Raw, log frequency of content words	-	+	-
Min raw, min log frequency of content words	-	+	-
Lexical variation: lexical, adjective, adverb, modifier, verb, noun	+	-	-
Lexical density	-	-	+
% lemmas belonging to the basic vocabulary	+	+	+
% lemmas belonging to the basic vocabulary and the fundamental usage repertoire, high usage repertoire, high availability	-	-	+
% token overlap with Elementary, Middle, High-school class	-	+	-
Mean hypernymy value of nouns, verbs	-	+	-
normal, logarithmic, and logarithmic per million words word frequency (based on Google Books 2012 and SUBTLEX-IT reference corpora)	+	-	-
informativeness per million words features (based on Google Books 2012 and SUBTLEX-IT reference corpora)	+	-	-
familiarity per million words features (based on Google Books 2012 and SUBTLEX-IT reference corpora)	+	-	-
logarithmic contextual diversity features (based on Google Books 2012 and SUBTLEX-IT reference corpora)	+	-	-

Imageability	+	-	-
Concreteness	+	-	-
Age of acquisition	+	-	-
Lexical readability index	-	-	+
Morpho-syntactic complexity indices			
% nouns	+	-	+
% singular, plural nouns	+	-	-
% proper nouns	+	-	+
% adjectives	+	-	+
% singular, plural adjectives	+	-	-
% possessive adjectives	+	-	-
% verbs	+	-	+
% conjunctions	+	-	+
% coord. conjunctions	+	-	+
% subord. conjunctions	+	-	+
% personal pronouns	+	+	-
% possessive, indefinite, relative, interrogative, demonstrative pronouns	+	-	-
% adverbs, negation adverbs	+	-	-
% prepositions	+	-	-
% main, finite, non-finite, infinite, gerund, 3rd person, 3rd person singular, indicative, indicative future, indicative past, indicative imperfect, imperative, conditional present, conjunctive, conjunctive present, conjunctive imperfect, modal, auxiliary verb	+	-	-
% ordinal, cardinal numbers	+	-	-
% interjection	+	-	-
% abbreviation	+	-	-
% twitter tag	+	-	-
% emoticon	+	-	-
% symbols	+	-	-
% punctuation	+	-	-
% modifier	+	-	-
% functional words	+	-	-
% articles	+	-	-
% foreign words	+	-	-
Syntactic complexity indices			
Number of clauses per sentence	-	-	+
% of main clauses	-	-	+
% of subordinate clauses	-	-	+
adjectival clause modifiers, adjectival modifiers, adverbial clauses, adverbial modifiers, appositional modifiers, attributives, auxiliaries, auxiliary passives, coordinations, subordinate clauses per sentence	+	-	-
Mean number of tokens per clause	-	-	+
Mean number of dependents per verbal head	-	-	+
Mean depth of complex noun structures	-	-	+
Mean maximum syntactic tree depth	-	-	+
Mean depth of subordinate chains	-	-	+
Mean length of dependency relations (distance in tokens from head to dependent)	-	-	+
Mean of maximum lengths of dependency relations (distance in tokens from head to dependent)	-	-	+
Noun phrase incidence	-	+	-
Mean noun modifiers per noun phrase	-	+	-
Higher level constituents	-	+	-
Mean number of tokens before main verb	-	+	-
Number of connectives and number of connectives per token	+	-	-
Causal cohesion	+	+	-
Syntactic similarity of all sentences	-	+	-
Causal content	-	+	-
Incidence of positive and negative additive, logical, temporal, and causal connectives	-	+	-
Incidence of all connectives	+	+	-
Incidence of conditional operators	-	+	-
Intentional content	-	+	-
Intentional cohesion	-	+	-
Temporal cohesion	+	+	-
Spacial cohesion	-	+	-
Number of syntactic constituents	+	-	-
Syntactic readability index	-	-	+
Readability indices			
Basic readability index	-	-	+
Global readability index	-	-	+
Gulpease	+	+	+

Table 2: Complexity measures implemented in CTAP, Coease and READ-IT. '+' indicates that a certain measure is implemented in this particular tool and '–' indicates that it is not implemented.

8. Bibliographical References

Alexopoulou, T., Michel, M., Murakami, A., and Meurers, D. (2017). Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing natural language processing techniques, *Language Learning*, 67:181–209.

Aluisio, S., Specia, L., Gasperin, C., Scarton, C. (2010). Readability Assessment for Text Simplification, Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications, pp. 1–9, Los Angeles, California, June.

Aprosio, A. P. and Moretti, G. (2018). Tint 2.0: An All-inclusive Suite for NLP in Italian, Proceedings of the Fifth Italian Conference on Computational Linguistics (CliC-it 2018).

Aprosio, A. P. and Moretti, G. (2016). Italy goes to Stanford: a collection of CoreNLP modules for Italian, *ArXiv e-prints*: http://adsabs.harvard.edu/abs /2016arXiv 160906204P

Brezina, V. and Pallotti, G. (2019). Morphological complexity in written l2 texts, *Second Language Research*, 35(1):99–119.

Bulté, B. and Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity, *Journal of Second Language Writing*, 26(0):42–65.

Burani, C., Arduino, L. S., Barca, L. (2001). Una base di dati sui valori di età di acquisizione, frequenza, familiarità, immaginabilità, concretezza, e altre variabili lessicali e sublessicali per 626 nomi dell'italiano, *Giornale Italiano di Psicologia*, January 2001.

Chen, X. and Meurers, D. (2016). CTAP: A Web-Based Tool Supporting Automatic Complexity Analysis, Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity, Osaka, Japan, 11-17 December 2016, pp. 113–119.

Chen, X., and Meurers, D. (2018). Word frequency and readability: Predicting the text-level readability with a lexical-level attribute, *Journal of Research in Reading, 41*(3):486–510.

Collins-Thompson, K. (2014). Computational assessment of text readability: A survey of current and future research, *ITL-International Journal of Applied Linguistics*, John Benjamins, 165(2):97–135.

Crepaldi, D., Amenta, S., Mandera, P., Keuleers, E., Brysbaert, M. (2015). SUBTLEX-IT. Subtitle-based Word Frequency Estimates for Italian, Annual Meeting of the Italian Association for Experimental Psychology, Rovereto, 10-12 September 2015.

Crossley, S., Weston., J., McLain Sullivan, S., and McNamara, D. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis, *Written Communication*, 28(3):282–311.

Crossley, S. and McNamara, D. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners, *Journal of Second Language Writing*, 26:66–79.

De Clercq, B., and Housen, A. (2019). The development of morphological complexity: A cross-linguistic study of L2 French and English, *Second Language Research Special Issue on Linguistic Complexity*, 35(1):71–97.

De Mauro, T., a cura di (2016). Il Nuovo vocabolario di base della lingua italiana, 23 dicembre 2016, https://dizionario.internazionale.it/nuovovocabolariodibase

Dell'Orletta, F. (2009). Ensemble system for Part-of-Speech tagging, Proceedings of Evalita'09, Evaluation of NLP and Speech Tools for Italian, Reggio Emilia, December 2009.

Dell'Orletta, F., Montemagni, S., Venturi, G. (2011). Assessing Readability of Italian Texts with a View to Text Simplification, Proceedings of the 2nd Workshop on Speech and Language Processing for Assistive Technologies, pp. 73–83.

Dell'Orletta, F., Montemagni, S., Venturi, G. (2013). Linguistic profiling of texts across textual genres and readability levels. An exploratory study on Italian fictional prose, Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013, pp. 189–197.

DuBay, W. H. (2004). The Principles of Readability. Impact Information, Costa Mesa, California.

Feng, L., Jansche, M., Huenerfauth, M., and Elhadad, N. (2010). A comparison of features for automatic readability assessment, Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010), pp. 276–284, Beijing, China.

Ferrucci, D. and Lally, A. (2004). UIMA: An Architectural Approach to Understanding Information Processing in the Corporate Research Environment, *Natural Language Engineering*, 10(3-4), September 2004.

Flesch, R. (1948). A new readability yardstick, *Journal of applied psychology*, 32 (3):221–233.

Franchina, V. and Vacca, R. (1986). Adaptation of Flesh readability index on a bilingual text written by the same author both in Italian and English languages, *Linguaggi*, 3:47–49.

François, T. and Miltsakaki, E. (2012). Do NLP and machine learning improve traditional readability formulas? *Applied natural language processing: Identification, investigation and resolution*, IGI Global, pp. 188–205.

Housen, A., and Kuiken, F. (2009). Complexity, accuracy and fluency in second language acquisition, *Applied Linguistics*, 30(4):461–473.

Housen, A., De Clercq, B., Kuiken, F., and Vedder, I. (2019). Multiple approaches to complexity in second language research, *Second Language Research Special Issue on Linguistic Complexity*, 35(1):2–31.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., Chissom, B. S. (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*, Millington, TN: Naval Technical Training, U. S. Naval Air Station, Memphis, TN.

Kyle, K. (2016). Measuring Syntactic Development in L2 Writing: Fine Grained Indices of Syntactic Complexity and Usage-Based Indices of Syntactic Sophistication. Ph.D. thesis, Georgia State University.

Lucisano, P., Piemontese, M. E. (1988). Gulpease. Una formula per la predizione della difficoltà dei testi in lingua italiana, *Scuola e Città*, 3:57–68.

McCarthy, P. M., Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment, *Behavior Research Methods*, 42 (2):381–392.

McNamara, D. S. and Graesser, A. C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing, *Applied natural language processing: Identification, investigation and resolution*, IGI Global, pp. 188–205.

Lin, Y., Michel, J.-B., Aiden, E. L., Orwant, J., Brockman, W., Petrov, S. (2012). Syntactic Annotations for the Google Books Ngram Corpus, Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, Jeju, Republic of Korea, 8-14 July 2012, pp. 169–174.

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

Paquot, M. (2019). The phraseological dimension in interlanguage complexity research, *Second Language Research*, 35(1):121–145.

Pilán, I., Vajjala, S., Volodina, E. (2016). A readable read: Automatic assessment of language learning materials based on linguistic complexity, *International Journal of*

Computational Linguistics and Applications, 7(1):143–159.

Reynolds, R. (2016). Russian natural language processing for computer-assisted language learning: capturing the benefits of deep morphological analysis in real-life applications. Ph.D. thesis, UiT - The Arctic University of Norway.

Talignani, G. (2019). Bahamas, donna salva 100 cani dall'uragano ospitandoli nel suo appartamento, *La Repubblica*, September 3, 2019.

Tonelli, S., Manh, K. T., Pianta, E. (2012). Making Readability Indices Readable, *NAACL-HLT 2012* Workshop on Predicting and Improving Text Readability for target reader populations (PITR 2012), Montreal, Canada, June 7, 2012, pp. 40–48.

Vajjala, S. and Meurers, D. (2012). On improving the accuracy of readability classification using insights from second language acquisition, Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications (BEA), pp. 163–173, Montréal, Canada. ACL.

Vajjala, S. (2018). Automated assessment of non-native learner essays: Investigating the role of linguistic features. *International Journal of Artificial Intelligence in Education*, 28(1):79–105.

Venturi, G., Bellandi, T., Dell'Orletta, F., Montemagni, S. (2015). NLP-Based Readability Assessment of Health-Related Texts: A Case Study on Italian Informed Consent Forms, Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis, pp. 131–141.

Vögelin, C., Jansen, T., Keller, S., Machts, N., and Möller, J. (2019). The influence of lexical features on teacher judgements of ESL argumentative essays, *Assessing Writing*, 39:50–63.

Weiss, Z. and Meurers, D. (2018). Modeling the Readability of German Targeting Adults and Children: An empirically broad analysis and its cross-corpus validation, Proceedings of the 27th International Conference on Computational Linguistics (COLING), pp. 303–317.

Weiss, Z. and Meurers, D. (2019a). Broad Linguistic Modelling is Beneficial for German L2 Proficiency Assessment. In Abel, A., Glaznieks, A., Lyding, V., Nicolas, L. (eds.), *Widening the Scope of Learner Corpus Research, Selected Papers from the Fourth Learner Corpus Research Conference,* Louvain-la-Neuve: Presses Universitaires de Louvain, pp. 419–435.

Weiss, Z. and Meurers, D. (2019b): Analyzing Linguistic Complexity and Accuracy in Academic Language Development of German across Elementary and Secondary School, Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications (BEA), pp. 380–393, Florence, Italy.

Weiss, Z., Riemenschneider, A., Schröter, P., and Meurers, D. (2019). Computationally modelling the impact of task-appropriate language complexity and accuracy on human grading of German essays, Proceedings of the 14th Workshop on Innovative Use of NLP for Building Educational Applications (BEA), pp. 30–45, Florence, Italy.

Yoon, H-J. (2017). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality, *System*, 66:130–141.