

# A NLP-based Analysis of Reflective Writings by Italian Teachers

Giulia Chiriatti\*, Valentina Della Gala\*, Felice Dell’Orletta<sup>◇</sup>, Simonetta Montemagni<sup>◇</sup>,  
Maria Chiara Pettenati\*, Maria Teresa Sagri\*, Giulia Venturi<sup>◇</sup>

\*Università di Pisa

giuliachiriatti@gmail.com

\*Istituto Nazionale Documentazione, Innovazione, Ricerca Educativa (INDIRE)

{v.dellagala, mc.pettenati, t.sagri}@indire.it

<sup>◇</sup>Istituto di Linguistica Computazionale “Antonio Zampolli” (ILC-CNR) - ItaliaNLP Lab

{nome.cognome}@ilc.cnr.it

## Abstract

**English.** This paper reports first results of a wider study devoted to exploit the potentialities of a NLP-based approach to the analysis of a corpus of reflective writings on teaching activities. We investigate how a wide set of linguistic features allows reconstructing the linguistic profile of the texts written by the Italian teachers and predicting whether are reflective.

**Italiano.** L’articolo descrive i primi risultati di uno studio più ampio che impiega strumenti e metodi di analisi e classificazione automatica del testo per descrivere le caratteristiche linguistiche di un corpus di documenti scritti dai neoassunti nella scuola italiana che riflettono su una specifica esperienza didattica.

## 1 Introduction

Since 2014, the “National Institute for Documentation, Innovation and Educational Research” (INDIRE) manages for the Ministry of Education (MIUR) the *induction program* of the Italian Newly Qualified Teachers (NQTs), i.e. the induction phase of teachers professional development that aims to support teachers in their transition from their initial teacher education into working life in schools. Experimented for the first time in 2014, it became effective starting in 2015 with the DM 850/2015.<sup>1</sup> The program involves all new hiring teachers from primary to secondary school for a total of 130,000 NQTs committed in the last 3 years. The underlying theoretical framework developed by INDIRE, MIUR and University of

Macerata is based on the alternation of laboratory and traditional classroom activities with documentation and reflection activities. The purpose is “to influence practices through a process that alternates between moments of immersion and distancing, which are actualised in *When I teach* and *When I reconsider my teaching to think of what happened*” (Magnoler et al., 2016). An on-line environment developed and managed by INDIRE<sup>2</sup> was set up to support teachers to reflect about and document their educational and professional activities (see Figure 1) during the induction program. All evidences of the instructional tasks (surveys, writing tasks, lesson plans, instructional materials, etc.) are collected in the e-portfolio and printed by the teachers for the final exam. An yearly monitoring of teachers activities is carried on by INDIRE to assess the effectiveness of the whole induction program, as well as of the single instructional tasks. It is aimed to modify, whenever needed, the program in order to improve stakeholders’ scaffolding to the newly qualified teachers and lastly teachers’ professional development.



Figure 1: The on-line environment collecting the e-portfolio of the newly qualified teachers.

In this paper, we report first results of an ongoing study devoted to investigate the potentialities offered by Natural Language Processing methods and tools for the analysis of the NQTs e-

<sup>1</sup>[http://neoassunti.indire.it/2018/files/DM\\_850\\_27\\_10\\_2015.pdf](http://neoassunti.indire.it/2018/files/DM_850_27_10_2015.pdf)

<sup>2</sup>The e-portfolio is available at <http://neoassunti.indire.it/2018/>

portfolio. We consider in particular the documents written by the 26,526 teachers hired in the 2016/17 school year. Many protocols (or models) have been proposed to assess reflection in teachers writing, e.g. (Sparks-Langer et al., 1990; Hatton and Smith, 1995; Kember et al., 2008; Larrivee, 2008; Harland and Wondra, 2011). These models rely on features that suggest either different levels of reflection (means focused on the depth of reflection) or content of reflection (focused on the breadth of reflection), and usually they have found to mix features of both classes (depth and breadth) (Ullmann, 2015). We rather focus here on the analysis of the *form* to study which are the main linguistic phenomena, distinguishing reflective from non reflective writings. Specifically, we devised a methodology devoted to investigate whether and to which extent a wide set of linguistic features automatically extracted from texts can be exploited to characterize NQTs' reflective writings.

**Our contribution:** *i*) we collect a corpus of reflective writings manually annotated by experts in the learning science domain and classified with respect to different types of reflectivity; *ii*) we detect a wide set of linguistically phenomena, characterizing the collected writings; *iii*) we report the first results of an automatic classification experiment to assess which features contribute more in the automatic prediction of reflexivity.

## 2 Defining reflection

Within the teaching and teacher education domain, a very large amount of studies have been dedicated to conceptualization and analysis of teachers reflection and teachers' reflective practice. Dewey (1933), Van Manen (1977), Schon (1984; Schon (1987; Schon (1991), Mezirow (1990) are among the main references. The attention on reflective thinking in the teachers education field has increased starting from the 80s as a reaction to the overlay technical view of teaching. Scholars have intensely studied reflection as a concept, detected more levels and types of reflection, how it works during and after professional teachers' practice, its role and purpose in teachers' professional development, and how it can be embedded in the curriculum of teachers preparation or professional development, and which techniques may be used to promote it (groups of discussion, readings, oral interview, action research projects, writing tasks, etc). In his seminal work "How we

think", Dewey provides the most shared definition of reflective thinking as applied in the educational field: reflection may be seen as an "active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusions to which tends". Hence, reflection is a systematic process of thinking that happens only if related to actual experiences, and includes observation of conditions and references to different pieces of knowledge, (i.e. references to previous experiences, domain knowledge, common sense knowledge, etc.), in order to respond to a dilemma (Mezirow, 1990). Teachers' educators have extensively employed writing tasks, such as writing structured or unstructured journals, portfolios, essays, blogs, open-ended questions to foster reflection both in pre-service and experienced teachers. Operational definitions of reflectivity proposed to develop schemes for assessing it are focused on identifying the presence of "reflective content" in teachers' writing, or how deep the reflection is.

Based on these premises, we are currently developing a reflection assessment schema suitable to describe properly the peculiarities of the Italian teachers' reflective writings written in the framework of the 2016/17 induction program. The schema designed so far, reported in Table 1, was devised according to the following criteria: a writing is reflective if it i) makes direct references to experienced teaching activity, ii) involves several topics (content/pedagogical knowledge) and references to previous experiences, classroom management, learners needs, iii) includes premises analysis (theoretical, context-related, personal) iv) debates a problem (a dilemma), a doubt, v) has an output: it sums up what was learned, sketches future plans, gives a new insight and understanding for immediate or future actions.

## 3 The Corpus

The corpus of NQTs reflective writings is part of the wider collection of documents written by the 26,526 teachers engaged in the 2016/17 INDIRE induction program. The whole corpus includes all texts written in two of the seven activities of the e-portfolio: *Didactic Activity 1* and *2* (DA) for a total of 265,200 texts. During these two activities, teachers were supported by guiding questions designed by INDIRE experts to help them to understand the consistency of the planned and acted

Type of reflectivity	Description	Example
No reflection	Simple writing that merely describes what happened during the teaching activity, no doubts or clues of an inquiry attitude are shown	I contenuti presentati sono stati acquisiti e gli alunni intervistati si sono dimostrati soddisfatti dell'intervento e del parere personale che hanno potuto esprimere sull'argomento di discussione.
General considerations and understanding	Writing shows weak links to the actual teaching experience, it is conducted at a distance from the phenomena of interest. It can include general thoughts and considerations	Per rispondere alla domanda circa la possibilità di migliorare l'attività affrontata, dirò innanzitutto che ritengo sempre possibile migliorare le proprie prestazioni. Sono convinta che l'esperienza sia una grande alleata e che, col tempo, si cresca, ci si arricchisca e si migliori.
Descriptive reflection	Writing includes considerations on actual classroom actions/events and some kind of knowledge base but doesn't clearly refer to any "problems", doubt or dilemma	Credo che la scelta più efficace sia stata quella della valutazione tra pari. In particolare, durante la fase della premiazione del concorso di poesia, un alunno per classe si è recato nell'altra scuola e ha tenuto un discorso introduttivo alla premiazione, nonché gestito la stessa in autonomia. Questo, a mio avviso, ha fatto sentire gli studenti i veri protagonisti del loro lavoro e ha favorito la motivazione, intrinseca ed estrinseca. Le consegne sono sempre state fornite in modo chiaro, ma hanno necessitato diverse ripetizioni per essere assimilate.
Reflection	Writing discusses problems, doubts and refers to some kind of action. It may report a reflective practice. There could be evidences of a change on teachers' attitude or acquiring new insights due to the problems faced	In realtà, mi sono accorta che solo pochi di loro erano capaci di dare una spiegazione adeguata (anche dal punto di vista formale) e soprattutto non riuscivano a trovare esempi calzanti se non con l'aiuto del libro di testo. Questo momento di ricognizione ha portato via quasi il doppio del tempo che avevo previsto, ma è comunque stato molto utile per accelerare il loro compito di ricerca durante l'analisi del nuovo testo proposto. Li ho stimolati a chiarire ogni dubbio e grazie anche alle loro domande credo che gli argomenti siano stati davvero appresi da tutti gli studenti, anche da chi di solito ha più difficoltà o da chi normalmente partecipa meno. È stata una lezione che li ha molto coinvolti nonostante si trattasse di una lezione piuttosto "tradizionale", perché mi hanno detto che questo sarebbe servito loro anche per lo studio di altre materie e soprattutto in vista dell'esame.

Table 1: Annotation schema of reflectivity.

teaching activities. For DA 1 and 2 they wrote 5 short texts as answers to 5 different groups of questions. The first 4 groups provide guidance for teachers to write general reflections only on the *design* of their teaching activity; the fifth group is meant to guide NQTs towards an overall reflection on their whole teaching experience, i.e. both the design and the real teaching activity, also including classroom assessment techniques.

We focused here on the answers to this latter group of questions that were devised in order to encourage teachers to reflect on the following issues: *i*) differences and similarities between the designed and achieved activities, *ii*) the most effective choices adopted, also including classroom assessment techniques, *iii*) how the activity could be improved, *iv*) the role played by the tutor and documentation practices. We considered in particular a subset of this group of answers that were annotated by 3 experts in the learning sci-

ence domain according to the reflectivity annotation scheme described in Section 2 (see Table 2). The agreement between the three annotators was calculated using the Fleiss' kappa test and we obtained a  $k=0.66$ , i.e. substantial agreement.

Reflectivity	n. answers	n. sent.	n. tokens
No reflection	185	348	9,784
Rhetoric	35	91	3,140
Reflection	217	609	21,686
Radical reflection	36	149	5,326
<b>TOTAL</b>	<b>473</b>	<b>1,197</b>	<b>39,936</b>

Table 2: Corpus of NQTs reflective writings annotated for different types of reflectivity.

#### 4 Linguistic Features and Reflectivity

The annotated corpus was tagged by the part-of-speech tagger described in Dell'Orletta (2009) and dependency-parsed by the DeSR parser (Attardi

et al., 2009). This allowed to extract a wide set of multilevel features, i.e. raw text, lexical, morpho-syntactic and syntactic, fully described by Dell’Orletta et al. (2013). They were used to reconstruct the linguistic profile of reflective writings and to carry out a first classification experiment aimed at predicting whether a text is reflective.

#### 4.1 Distribution of Linguistic Features

Table 3 shows a selection of the features that vary significantly *i*) between reflective and non-reflective answers (column *Reflectivity*) and *ii*) among the different types of reflectivity we considered (column *Types of Reflectivity*)<sup>3</sup>. The analysis of variance was computed in the first case using the Wilcoxon Rank-sum test for paired samples, while in the second case we used the Kruskal-Wallis test since we aimed to assess the different distribution of features in the 4 classes.

In both cases, features from all levels of analysis resulted to be significant. If we consider the first ten most discriminative features, reflective writings resulted to be longer in terms of number of words and sentences, they are characterized by longer sentences and by a lower Type/Token Ratio; they contain an higher number of verbal heads and of embedded complement ‘chains’ (governed by a nominal head). Interestingly, they mostly contain linguistic phenomena typically related to syntactic complexity, for example they are characterized by *i*) an higher use of verbal modification (e.g. higher % of adverbs, of auxiliary and modal verbs), *ii*) more complex verbal predicate structures (e.g. higher average verbal arity, calculated as the number of instantiated dependency links sharing the same verbal head), *iii*) more extensive use of subordination (e.g. higher % of subordinate clauses also embedded in deep chains), *iv*) features related to a non canonical word order (e.g. higher % of pre-verbal objects and post-verbal subjects), *v*) longer dependency links and higher parse trees, two features related to sentence length. On the contrary, non reflective NQTs’ answers contain an higher level of lexical complexity: they have an higher Type/Token Ratio, a lower percentage of “Fundamental words”, i.e. very frequent words according to the classification proposed by De Mauro (2000) in the *Basic Italian Vocabulary* (BIV), and an higher percentage of “High usage words”.

<sup>3</sup>The full list of ranked features is contained in Appendix.

If we focus on the linguistic profile of the different types of reflective writings, we can observe that answers annotated as *Reflection* and *Radical reflection* are mostly characterized by features typically related to structural complexity. This is particular the case of *Radical reflection* answers that are longer in terms of number of sentences and words; they have more complex verbal predicates (e.g. an higher % of adverbs and of an implicit mood such as gerundive that can be more ambiguous with respect to the referential subject), more complex use of subordination (e.g. average length of ‘chains’ of embedded subordinate clauses), long distance constructions (length of dependency links), non canonical constructions (post-verbal subject). The higher % of demonstrative pronouns and determiners can be related to one of the most representative characteristic of reflection, i.e. the direct reference to real life. On the contrary, they contain a simpler use of lexicon, e.g. a lower Type/Token ratio and an higher percentage of “Fundamental words”.

#### 4.2 Prediction of Reflectivity

Table 4 reports the results of the automatic classification experiment we devised in order to predict whether a text is reflective. We built a classifier based on LIBLINEAR (Fan et al., 2008) as machine learning library trained using the LIBLINEAR L2-regularized L2-loss support vector classification function. We followed a 5-fold cross-validation process and relied on a training set of 370 answers balanced between the reflective and non reflective texts, since the under sampling technique has been proofed to improve classification performance on unbalanced datasets (Qazi and Raza, 2012). The performance was calculated in terms of F-score in the correct classification of *non reflective* (0 in the table) or of *reflective* (1) writings. We used different classification models: the *Raw text* one uses only raw text features, the *Lexical* one uses the distribution of the lexicon belonging to the *Basic Italian Vocabulary* and up to bi-grams of words, the *Morpho-syntactic* one uses the unigram of part-of-speech and verbal morphology features, the *All features* model uses all the considered features including the syntactic ones. A very competitive baseline was computed: it exploits the distribution of unigrams of words (*Unigrams*). As it can be seen, the model that uses all the considered features resulted to be the best

Feature	Ranking position		Avg. Feature Value in different types of (non)reflective texts			
	Reflectivity	Types of Reflectivity	No reflection	Rhetoric	Reflection	Radical reflection
<b>Raw text features:</b>						
Avg sentence length	<b>10</b>	<b>11</b>	27.97	35.9	<b>38.6</b>	38.2
Avg number of sentences	<b>9</b>	<b>7</b>	1.88	2.6	2.81	<b>4.14</b>
Avg number of words	<b>1</b>	<b>1</b>	52.89	89.71	99.94	<b>147.94</b>
<b>Lexical features:</b>						
Type/token ratio (100 token)	<b>8</b>	<b>9</b>	<b>0.78</b>	0.71	0.7	0.69
% of “Fundamental words” of <i>BIV</i>	62	86	74.15	75.57	77.01	<b>77.92</b>
% of “High usage words” of <i>BIV</i>	92	<b>38</b>	<b>19.35</b>	15.79	15.71	14.92
% of “High availability words” of <i>BIV</i>	58	68	9.72	<b>12.8</b>	10.78	10.69
<b>Morpho-syntactic features:</b>						
% of adjectives	71	87	7.29	<b>9.16</b>	7.72	7.93
% of possessive adjectives	67	<b>43</b>	1.08	<b>2</b>	0.97	0.93
% of adverbs	<b>42</b>	<b>46</b>	3.95	3.93	4.82	<b>5.29</b>
% of prepositions	51	82	15.11	<b>17.08</b>	16.61	16.05
% of demonstrative pronouns	<b>36</b>	<b>34</b>	0.43	0.65	0.58	<b>0.78</b>
% of demonstrative determiners	<b>35</b>	<b>30</b>	0.35	0.66	0.42	<b>0.6</b>
% of determinative articles	<b>30</b>	<b>41</b>	<b>8.29</b>	6.89	6.81	7.07
% of subordinative conjunctions	69	63	0.94	0.68	0.98	<b>1.27</b>
% of sentence boundary punctuation	<b>12</b>	<b>12</b>	<b>4.17</b>	2.99	2.86	2.92
% of auxiliary verbs	<b>25</b>	<b>27</b>	<b>6.66</b>	4.01	4.92	4.48
% of modal verbs	<b>40</b>	<b>40</b>	0.69	<b>1.06</b>	0.78	0.97
% of verbs – subjective mood	72	<b>39</b>	1.16	1.29	<b>2.55</b>	1.53
% of verbs – infinitive mood	<b>28</b>	<b>36</b>	19.11	<b>27.48</b>	25.03	25.75
% of verbs – gerundive mood	<b>37</b>	<b>45</b>	5.54	6.06	6.51	<b>6.73</b>
% of verbs – indicative mood	<b>38</b>	58	10.46	<b>14.76</b>	11.74	12.91
% of verbs – third person singular	<b>20</b>	<b>15</b>	8.2	18.76	14.92	<b>19.3</b>
% of verbs – third person plural	80	91	6.14	<b>10.83</b>	8.04	7.67
% of verbs – imperfect tense	78	<b>35</b>	7.18	1.55	9.72	<b>13.75</b>
<b>Syntactic features:</b>						
% of dependency types – auxiliary	<b>24</b>	<b>25</b>	<b>6.65</b>	3.98	4.88	4.41
% of dependency types – object	44	59	4.22	4.7	5.06	<b>5.6</b>
% of dependency types – preposition	55	81	15.15	<b>17.33</b>	16.6	16.09
% of dependency types – subordinate clause	60	62	0.99	0.78	1.03	<b>1.22</b>
% of dependency types – subject	46	83	<b>4.62</b>	3.62	3.77	3.74
Avg number of verbal heads	<b>2</b>	<b>3</b>	52.89	89.71	99.94	<b>147.94</b>
Avg number of embedded complement chains	<b>4</b>	<b>4</b>	9.72	<b>12.8</b>	10.78	10.69
Length of ‘chains’ of embedded subordinate clauses (avg)	<b>19</b>	<b>21</b>	0.48	0.69	0.86	<b>0.95</b>
Maximum length of dependency links (avg)	<b>16</b>	<b>19</b>	10.26	12.71	14.16	<b>14.8</b>
Parse tree depth (avg)	<b>21</b>	<b>24</b>	7.86	9.73	9.56	<b>9.65</b>
Arity of verbal predicates (avg)	<b>13</b>	<b>13</b>	3.62	4.46	<b>4.89</b>	4.74
% of pre-verbal objects	52	<b>42</b>	4.84	<b>9.71</b>	7.59	4.81
% of post-verbal subject	86	84	10.65	11.17	10.64	<b>17.07</b>
% of subordinate clauses in post-verbal position	<b>23</b>	<b>16</b>	52.21	76.57	78.97	<b>97.71</b>

Table 3: Feature ranking position characterizing *i*) reflective vs. non reflective texts and *ii*) different types of reflective texts and average value of feature distribution in the different types of reflective texts. Ranking positions with  $p < 0.001$  are marked in italics and with  $p < 0.05$  in boldface.

one. On the contrary, the model relying on very simple types of features (raw text features) that capture how much teachers have written achieves the worst results. We also carried out a very preliminary experiment to classify the three different types of reflective writings but it produced unsatisfactory results due to the unbalanced distribution of answers in the reflective classes. As expected, a balanced experiment yielded very low accuracies since we used very few data.

## 5 Conclusions and current developments

We reported first results of a on-going study devoted to reconstruct the linguistic profile of a corpus of reflective writings by Italian newly recruited teachers that we collected for the specific purpose of this paper. We are currently enlarging

Features	F1 0	F1 1	Tot F1
Raw text	58.4	69.86	64.13
Lexical	78.58	77.53	78.05
Morpho-syntactic	74.87	75.18	75.02
All features	<b>79.31</b>	<b>79.01</b>	<b>79.16</b>
Baseline (unigrams)	75.16	74.84	75.00

Table 4: Classification of reflective vs. non reflective writings using different models of features.

the corpus with new manually annotated data to improve the accuracy of the automatic classification of different types of reflectivity.

## References

G. Attardi, F. Dell’Orletta, M. Simi and J. Turian. 2009. Accurate dependency parsing with a stacked

- multilayer perceptron. *Proceedings of Evalita'09, Evaluation of NLP and Speech Tools for Italian*, Reggio Emilia, December.
- D. Boud and D. Walker. 2013. *Reflection: Turning Experience into Learning*. RoutledgeFalmer.
- C.C. Chang and C.J. Lin. 2001. LIBSVM: a library for support vector machines. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- F. Dell'Orletta. 2009. Ensemble system for Part-of-Speech tagging. *Proceedings of Evalita'09, Evaluation of NLP and Speech Tools for Italian*, Reggio Emilia, December.
- F. Dell'Orletta, S. Montemagni and G. Venturi. 2013. Linguistic profiling of texts across textual genre and readability level. An exploratory study on Italian fictional prose. *Proceedings of the Recent Advances in Natural Language Processing Conference (RANLP-2013)*.
- T. De Mauro. 2000. *Grande dizionario italiano dell'uso (GRADIT)*. Torino, UTET.
- J. Dewey. 1933. *How we think: a restatement of the relation of reflective thinking to the educative process*. D.C. Heath and company.
- R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X. Wang, and C.-J. Lin. 2008. LIBLINEAR: A Library for Large Linear Classification. *Journal of Machine Learning Research*, 9:1871–1874.
- DJ. Harland and JD. Wondra 2011. Preservice Teachers' Reflection on Clinical Experiences: A Comparison of Blog and Final Paper Assignments. *Journal of Digital Learning in Teacher Education*, Vol. 27(4).
- N. Hatton and D. Smith. 1995. Reflection in teacher education: Towards definition and implementation. *Teaching and Teacher Education*, Vol. 11(1).
- D. Kember, J. McKey, K. Sinclair, FKY Wong 2008. A four category scheme for coding and assessing the level of reflection in written work. *Assessment and Evaluation in Higher Education*, Vol. 25(4).
- B. Larrivee 2008. Development of a tool to assess teachers' level of reflective practice. *Reflective Practice*, Vol. 9(3).
- P. Magnoler, GR. Mangione, MC. Pettenati, A. Rosa, PG. Rossi. 2016. Induction models and teachers professional development. *Journal of e-Learning and Knowledge Society*, Vol. 12(3).
- J. Mezirow. 1990. *Fostering critical reflection in adulthood: a guide to transformative and emancipatory learning*. Jossey-Bass Publishers.
- N. Qazi and K. Raza. 2012. Effect of Feature Selection, SMOTE and under Sampling on Class Imbalance Classification. *Proceedings of the 2012 UKSim 14th International Conference on Modelling and Simulation*, pp. 145-150.
- D.A. Schon. 1984. *The Reflective Practitioner: How Professionals Think In Action*. Basic Books.
- D.A. Schon. 1987. *Educating the Reflective Practitioner*. Jossey-Bass.
- D.A. Schon. 1991. *The reflective turn: Case studies in and on educational practice*. Teachers College Press.
- GM. Sparks-Langer, GM. Simmons, M. Pasch, A. Colton, A. Starko. 1990. Reflective pedagogical thinking: How can we promote it and measure it? *Journal of Teacher Education*, Vol. 41(5).
- T. D. Ullmann. 2015. *Automated detection of reflection in texts. A machine learning based approach*. The Open University.
- T. D. Ullmann. 2015. *Keywords of written reflection - a comparison between reflective and descriptive datasets*. Proceedings of the 5th Workshop on Awareness and Reflection in Technology Enhanced Learning.
- M. Van Manen. 1977. Linking Ways of Knowing with Ways of Being Practical. *Curriculum Inquiry*, Vol. 6(3).
- H.C. Waxman et al. 1987. *Images of Reflection in Teacher Education*. Summaries of papers presented at a National Conference on Reflective Inquiry in Teacher Education, Houston.

Feature	Ranking position		Avg. Feature Value in different types of (non)reflective texts			
	Reflectivity	Types of Reflectivity	No reflection	Rhetoric	Reflection	Radical reflection
<b>Raw text features:</b>						
Avg sentence length	<b>10</b>	<b>11</b>	27.97	35.9	<b>38.6</b>	38.2
Avg number of sentences	<b>9</b>	<b>7</b>	1.88	2.6	2.81	<b>4.14</b>
Avg number of tokens	<b>1</b>	<b>1</b>	52.89	89.71	99.94	<b>147.94</b>
<b>Lexical features:</b>						
Type/token ratio (first 100 lemma)	<b>8</b>	<b>9</b>	<b>0.78</b>	0.71	0.7	0.69
Type/token ratio (first 200 lemma)	<i>6</i>	<i>6</i>	<b>0.77</b>	0.68	0.67	0.64
% of "Fundamental words" of <i>BIV</i>	<i>62</i>	<i>86</i>	74.15	75.57	77.01	<b>77.92</b>
% of "High usage words" of <i>BIV</i>	<i>92</i>	<b>38</b>	<b>19.35</b>	15.79	15.71	14.92
% of "High availability words" of <i>BIV</i>	<i>58</i>	<i>68</i>	9.72	<b>12.8</b>	10.78	10.69
<b>Morpho-syntactic features:</b>						
Lexical density	<b>64</b>	<b>96</b>	0.54	0.55	0.55	<b>0.56</b>
% of adjectives	<i>71</i>	<i>87</i>	7.29	<b>9.16</b>	7.72	7.93
% of possessive adjectives	<i>67</i>	<b>43</b>	1.08	<b>2</b>	0.97	0.93
% of adverbs	<b>42</b>	<b>46</b>	3.95	3.93	4.82	<b>5.29</b>
% of negative adverbs	<b>54</b>	<b>53</b>	0.64	0.38	0.64	<b>0.65</b>
% of determiners	<b>63</b>	<b>88</b>	1.19	1.19	1.28	<b>1.43</b>
% of demonstrative determiners	<b>35</b>	<b>30</b>	0.35	0.66	0.42	<b>0.6</b>
% of indefinite determiners	<b>74</b>	<b>71</b>	0.8	0.47	<b>0.83</b>	0.8
% of prepositions	<i>51</i>	<i>82</i>	15.11	<b>17.08</b>	16.61	16.05
% of articles	<i>93</i>	none	<b>9.36</b>	8.34	8.38	<b>8.64</b>
% of demonstrative pronouns	<b>36</b>	<b>34</b>	0.43	0.65	0.58	<b>0.78</b>
% of personal pronouns	<b>89</b>	<b>99</b>	0.29	<b>0.39</b>	0.32	0.24
% of relative pronouns	<i>39</i>	<b>56</b>	1.17	1.16	1.48	<b>1.55</b>
% of determinative articles	<b>30</b>	<b>41</b>	<b>8.29</b>	6.89	6.81	7.07
% of subordinative conjunctions	<i>69</i>	<i>63</i>	0.94	0.68	0.98	<b>1.27</b>
% of single commas or hyphens	<i>27</i>	<i>33</i>	3.55	4.7	4.67	<b>5.26</b>
% of numbers	<b>87</b>	<b>67</b>	0.22	0.19	<b>0.4</b>	0.29
% of sentence boundary punctuation	<b>12</b>	<b>12</b>	<b>4.17</b>	2.99	2.86	2.92
% of verbs	<b>48</b>	<b>70</b>	<b>20.51</b>	17.71	18.52	17.91
% of auxiliary verbs	<b>25</b>	<b>27</b>	<b>6.66</b>	4.01	<b>4.92</b>	4.48
% of modal verbs	<b>40</b>	<b>40</b>	0.69	<b>1.06</b>	0.78	0.97
% of verbs – subjective mood	<i>72</i>	<b>39</b>	1.16	1.29	<b>2.55</b>	1.53
% of verbs – infinitive mood	<b>28</b>	<b>36</b>	19.11	<b>27.48</b>	25.03	25.75
% of verbs – gerundive mood	<b>37</b>	<b>45</b>	5.54	6.06	6.51	<b>6.73</b>
% of verbs – indicative mood	<b>38</b>	<i>58</i>	10.46	<b>14.76</b>	11.74	12.91
% of verbs – third person singular	<b>20</b>	<b>15</b>	8.2	18.76	14.92	<b>19.3</b>
% of verbs – third person plural	<i>80</i>	<i>91</i>	6.14	<b>10.83</b>	8.04	7.67
% of verbs – imperfect tense	<i>78</i>	<b>35</b>	7.18	1.55	9.72	<b>13.75</b>
<b>Syntactic features:</b>						
% of syntactic roots	<i>14</i>	<i>14</i>	<b>4.57</b>	3.06	3.36	3.21
% of dep-auxiliary	<b>24</b>	<b>25</b>	<b>6.65</b>	3.98	4.88	4.41
% of dep-nominal/clausal argument	<b>61</b>	<b>98</b>	2.36	<b>3.08</b>	2.8	2.41
% of dep-indirect complement	<b>66</b>	<b>61</b>	0.46	0.62	0.5	0.48
% of dep-locative complement	<b>47</b>	<i>31</i>	0.07	0.21	<b>0.34</b>	0.14
% of dep-temporal complement	<i>41</i>	<i>28</i>	0.16	0.3	0.28	<b>0.41</b>
% of dep-nominal/clausal modifier	<b>45</b>	<b>73</b>	15.88	17.25	17.07	<b>17.7</b>
% of dep-relative modifier	<i>32</i>	<i>32</i>	1.18	1.1	1.46	<b>1.8</b>
% of dep-object	<i>44</i>	<i>59</i>	4.22	4.7	5.06	<b>5.6</b>
% of dep-preposition	<i>55</i>	<i>81</i>	15.15	<b>17.33</b>	16.6	16.09
% of dep-subordinate clause	<i>60</i>	<i>62</i>	0.99	0.78	1.03	<b>1.22</b>
% of dep-subject	<i>46</i>	<i>83</i>	<b>4.62</b>	3.62	3.77	3.74
Avg number of verbal heads	<b>2</b>	<b>3</b>	52.89	89.71	99.94	<b>147.94</b>
Avg number of embedded complement chains	<b>4</b>	<b>4</b>	9.72	<b>12.8</b>	10.78	10.69
Length of 'chains' of embedded subordinate clauses (avg)	<b>19</b>	<b>21</b>	0.48	0.69	0.86	<b>0.95</b>
Length of dependency links (avg)	<i>15</i>	<i>18</i>	2.09	2.3	2.4	<b>2.42</b>
Maximum length of dependency links (avg)	<b>16</b>	<b>19</b>	10.26	12.71	14.16	<b>14.8</b>
Parse tree depth (avg)	<b>21</b>	<b>24</b>	7.86	9.73	9.56	<b>9.65</b>
Arity of verbal predicates (avg)	<b>13</b>	<b>13</b>	3.62	4.46	<b>4.89</b>	4.74
% of verbal roots	<b>57</b>	<i>29</i>	<b>0.96</b>	<b>0.95</b>	0.9	0.84
% of verbal roots with explicit subj	<b>70</b>	<b>65</b>	67.92	<b>73.76</b>	59.05	60.79
% of finite complement clauses	<b>83</b>	<b>95</b>	19.85	17.19	23.08	<b>27.64</b>
% of infinite complement clauses						
% of pre-verbal objects	<i>52</i>	<b>42</b>	4.84	<b>9.71</b>	7.59	4.81
% of post-verbal subject	<i>86</i>	<i>84</i>	10.65	11.17	10.64	<b>17.07</b>
% of subordinate clauses in post-verbal position	<b>23</b>	<b>16</b>	52.21	76.57	78.97	<b>97.71</b>

Table 5: **Appendix A:** Full list of feature ranking positions characterizing *i)* reflective vs. non reflective texts and *ii)* different types of reflective texts and average value of feature distribution in the different types of reflective texts. Ranking positions with  $p < 0.001$  are marked in italics and with  $p < 0.05$  in boldface. Features which were not selected during ranking have no rank.