



# ITACA

## Corpus Handbook

Edited by

Arianna Bienati, Jennifer-Carmen Frey, Lorenzo Zanasi, Chiara Vettori

Institute for Applied Linguistics  
Eurac Research

Version 1.0  
December 2024

## Contents

Corpus figures and sample description .....	3
Population description and sampling procedures .....	3
Corpus metadata .....	4
Text and person-related metadata .....	4
Text and sample identifiers .....	4
Author metadata .....	5
Metadata on reading and writing habits .....	11
Reading attitudes .....	11
Writing attitudes .....	12
Writing habits .....	13
Reading habits .....	14
Reading habits - parents .....	20
Coherence ratings .....	21
References .....	24

## Appendices

- A. Annotation scheme
- B. Rating scale

## Corpus figures and sample description

The ITACA corpus comprises 635 texts, for which metadata about the writer and score on a language assessment test were collected. 14 students did not complete the questionnaire and 22 did not complete the language assessment test. Therefore, the students who completed both questionnaire and writing task are 621 and those who completed all tasks were 603.

## Population description and sampling procedures

For the purpose of analysis, a representative sample of the South Tyrolean students' population enrolled in schools with Italian as the medium of instruction in the school year 2021/2022, stratified per school type (vocational-technical vs. academic high school) was drawn from the entire corpus of 635 texts. These texts are marked with the metadata `sample_repr_south_tyrol=True`.

The sample was calculated considering the strata related to students enrolled in vocational-technical schools (*istituti*) and in academic high schools (*licei*). It adds up to a total of 568 students, of which 554 have completed both the writing task and the sociolinguistic questionnaire.

A total of 1308 students were enrolled in the fourth year of vocational-technical (*istituti tecnici e professionali*) and academic high schools (*licei*) in the Province of Bozen/Bolzano in the school year 2021/2022. Of these, 43.4% were enrolled in vocational-technical schools and 56.6% in academic high schools. A representative sample of this student population results in 588 valid participants (and texts) to be drawn from all educational institutions in South Tyrol. Therefore, accounting for potentially unusable texts with a coefficient known from previous research conducted in the territory, a sample of 659 students was drawn. Knowing the number of classes and the average class size per educational institution, 38 classes were extracted, respectively 18 classes in vocational-technical schools and 20 classes from academic high schools, to reflect the distribution of students enrolled in different types of schools. In the extraction, a criterion of proportionality was applied: if an institution covered the 8% of students enrolled in academic high schools, an appropriate number of classes from that institution was extracted to cover approximately 8% of the academic high school sample.

The sample to be drawn was made up as follows:

	<b>N students</b>	<b>%</b>	<b>588 sample</b>	<b>Corrective surcharge</b>	<b>Classes to extract</b>
Istituto	567	43.4%	255	286	18
Liceo	741	56.6%	333	373	20
	<b>1308</b>	<b>100.0%</b>	<b>588</b>	<b>659</b>	<b>38</b>

Starting from these calculations, educational institutions have been chosen for data collection. However, given the difficulties in convincing some schools to participate, the return on the type

of educational institution was as follows: 90% of the vocational-technical schools in the province participated with at least one class (9 out of 10 schools) and 82% of the academic high schools in the province participated with at least one class (9 out of 11 schools).

	<b>Planned sample</b>	<b>Actual sample</b>	<b>Sample coverage</b>
Istituti	10	9	90%
Licei	11	9	82%

Within those 9 *licei* and *istituti*, 18 and 20 classes have been chosen for data collection. From the texts collected in these classes, 554 students have been randomly chosen for creating a sample representative of the South Tyrolean school reality. This sample maintains almost perfectly the proportion between school types envisaged in the beginning and it has been thus deemed representative of the student population under investigation.

	<b>Collected sample</b>	<b>Sample proportions</b>	<b>Classes in the sample that could be used</b>
Istituto	248	44.8%	18
Liceo	306	55.2%	20
	<b>554</b>	<b>100.0%</b>	<b>38</b>

## Corpus metadata

### Text and person-related metadata

The following text and person-related metadata is available for the ITACA corpus can be found in `/metadata/core_metadata.csv` and used in the search interface.

### Text and sample identifiers

`text_id/Id`

Unique and anonymous identifier for each student essay composed of a combination of a six-character long anonymous identifier for the class and a six-character long identifier for the respective student that are joined with an underscore (e.g., AC20PA\_BAZ14M).

`sample_repr_south_tyrol`

Information, whether the essay was part of the statistically drawn representative sample. Note that `core_metadata.csv` contains reference to some essays that were collected but had to be filtered out for various reasons at a later stage. Texts pertaining to the representative sample are marked with `TRUE`, all other texts are marked with `FALSE`.

`sample_annotated`

Information on whether the essay was annotated for cohesion and coherence features according to the annotation schema described below. The sample consist of 388 texts. Of which 40 texts belong to the sample of curated texts, annotated by all three annotators and the remaining 348, which have been annotated by only one out of three annotators.

`sample_curated`

Information, whether the essay was part of the curated sample. Curated texts represent the ITACA gold standard. It comprises a set of 40 texts (10% circa of the whole annotated corpus), of which 10 are annotated by all three annotators who took part in the annotation process and the remaining 30 are annotated by couples of two annotators.

All these texts have been reviewed by the project team members during the curation phase, a moment in which discrepancies between annotations made by different annotators have been discussed and resolved to arrive at a consensus and build the gold standard.

`author_id`

Unique and anonymous identifier for each student, i.e. the author of the essay, composed of six alphanumeric characters.

`class_id`

Unique and anonymous identifier for the school class in which the essay has been collected, composed of six alphanumeric characters.

`school_id`

Unique and anonymous identifier for the school in which the essay has been collected, composed of 2 characters.

`school_type`

The type of school in which the essay was collected.

- `liceo`, i.e. academic high school
- `istituto`, i.e. vocational-technical high school

**Note:**

Each author contributed with one essay. Authors are nested in classes which are again nested in schools of different school types. Missing values are encoded with 'NA'.

## **Author metadata**

Person-related metadata on study participants have been collected in an online questionnaire and can be identified by the metadata prefix `author_*`. These include general sociodemographic information, metadata regarding the language background of the student as well as metadata on reading and writing habits (that are, however, recorded in the file `rw_habits.csv`).

## **Sociodemographic metadata on participating students**

The following items were collected in the project:

author\_birthyear

The year in which the student was born, elicited by the question: "In che anno sei nato/a? (quattro cifre)". Possible responses include 4-digit birthyears between 2001-2005.

author\_gender

The students self-defined gender, elicited by the question: "Sei un/a". Possible responses include: Maschio | Femmina | Preferisco non dirlo | Altro

author\_dsa

Encodes the presence of learning disabilities and special needs (Disturbi Specifici dell'Apprendimento), elicited by the questionnaire item: "Hai una diagnosi di DSA (ad es: dislessia, disgrafia, discalculia, ecc.)?". Possible responses include: Sì | No

author\_legge\_104

Encodes the presence of other disabilities defined in the Italian legislation (legge 104/1992), elicited by the questionnaire item: "Sei portatore/portatrice di handicap ai sensi della legge 104/1992?". Possible responses: Sì | No

author\_SES

The student's socio-economic background. This value was calculated as the first dimension of a factor analysis using questionnaire items on the student's home possessions, the parents' highest educational degree and the parents' job category (see section Calculation of socio-economic status variable).

The following questionnaire items have been regarded.

For the highest educational degree of the parents:

- 'Indica il titolo di studio più alto conseguito dai tuoi genitori. Seleziona l'opzione "non si applica" in assenza della figura genitoriale.'

For the highest job classification of the parents:

- 'I tuoi genitori lavorano? Seleziona l'opzione "Non si applica" in assenza della figura genitoriale.'
- 'Tra le seguenti categorie, indica quale descrive meglio la professione dei tuoi genitori. Se in pensione, non più abile al lavoro, in cassa integrazione o disoccupato/a, fai riferimento all'ultima professione svolta.'
- 'Puoi descrivere la professione dei tuoi genitori? Denominazione del lavoro (ad es. operaio tornitore qualificato, magazziniere, avvocatessa, insegnante di scuola secondaria di I grado, ragioniera in un ufficio amministrativo, ecc.). Se deceduto/a, in pensione, in cassa integrazione o disoccupato/a fai riferimento all'ultima professione svolta, mentre nel caso in cui la professione sia sconosciuta indica non lo so. Ti preghiamo di essere il più preciso/a possibile.'

For students' home possessions:

- 'A casa tua ci sono:'
  - Una camera per te
  - Un computer che puoi usare per lo studio
  - Un collegamento Internet
  - Libri di letteratura classica (ad es. Dante)

- Opere d'arte (ad es. quadri)
- Una lavastoviglie
- Un televisore a schermo piatto
- 'Quante di ciascuna delle seguenti cose ci sono a casa tua?'
  - Smartphone
  - Televisore/i
  - Computer/tablet
  - Automobile/i
  - Locali con vasca da bagno/doccia
- 'Quanti libri ci sono a casa tua? Considera che ogni metro di scaffale contiene circa 40 libri. Non includere nel calcolo le riviste, i giornali o libri scolastici.'
  - 0-10 libri
  - 11-25 libri
  - 26-100 libri
  - 101-200 libri
  - 201-500 libri
  - Più di 500 libri

### **Calculation of socio-economic status variable**

For the purposes of the study, we furthermore constructed an Economic, Social and Cultural Status index to measure the effect of family background on the learning outcomes of the students participating in the ITACA project. This research largely follows the methodology proposed by OECD-PISA (PISA, 2018), albeit with some modifications dictated by the specificity of the data taken into consideration.

The ESCS index is based on the following variables:

- parents' employment status (HISEI)
- parents' level of education expressed in years of formal education (PARED)
- possession of certain material goods understood as proximity variables of an economic-cultural context favourable to learning (HOMEPOS).

The information used for the construction of these variables was obtained from the sociolinguistic questionnaire administered to the students (see [Sociolinguistic questionnaire \(in Italian\)](#)), more specifically from the information gathered with the questions 40-50. The questionnaire data contained information for 621 of the 635 students. For the remaining 14 students, who did not fill in the questionnaire, ESCS could not be calculated resulting in a missing value.

#### *The employment status of parents (HISEI)*

The information used for the construction of the index of the occupational status of the student's parents (HISEI) was obtained from the questions 40-42 (with respect to parent 1) and the questions 43-45 (with respect to parent 2) of the socio-demographic part of the questionnaire administered to the pupils. The occupational data of the student's parent 1 and parent 2 were obtained from the answers to the open-ended questions (Q42 and Q45). The answers were coded with the four-digit ISCO codes (International Labour Organization, n.d.) and then mapped to the International Socioeconomic Employment Status Index (ISEI) using the R package ISCO08ConveRsions (Schwitter, 2023). Based on this information, three indices

were calculated: the employment status of parent 1; the employment status of parent 2; and the highest parental employment status (HISEI), which corresponds to the highest ISEI score of either parent or the ISEI score of the only available parent. For all three indices, higher ISEI scores indicate higher levels of employment status. Information on the occupation of both parents were missing in 57 students of the 635 students in the ITACA corpus.

#### *Parents' level of education (PARED)*

Another variable related to the student's family background that was taken into account was the parents' level of education. Based on the information given in Q39 which asked for the educational qualification of both parents (codified in the seven ISCED levels), the PARED index provides an estimate of years of schooling according to the highest educational qualification of both parents. In 53 cases the information on the educational qualification was missing for both parents and could thus not be calculated.

#### *Economic-cultural well-being (HOMEPOS)*

In this study, economic wellbeing was explored through information derived from questions 49 and 50, in which students reported the quantity of certain goods (Q49) and the estimated number of books present in the home (Q50). In this case, data was missing just for those students who did not fill in the questionnaire (N=14). The synthetic variable HOMEPOS was calculated using the R package TAM which offers various types of Item Response Models. We based our HOMEPOS variable on the weighted likelihood estimates of a Partial Credit Model (PCM) (Masters & Wright, 1997). The PCM model represents a form of Rasch Model for polynomial data.

#### *The Economic, Social and Cultural Status index (ESCS)*

This index has been defined and consequently calculated following the methodological indications that have found greater acceptance in the international context. The ESCS was constructed on the basis of the three indices illustrated above: the parents' employment status (HISEI), their level of education (PARED) and whether or not they own particular capital and cultural assets (HOMEPOS).

The ESCS composite index is derived by means of a principal component analysis (PCA) of the three indices HISEI, PARED and HOMEPOS. As a preliminary step to the PCA, missing values were imputed with the R package missMDA (Josse & Husson, 2016). The results of the analysis report that the first extracted component is able to explain approximately 64% of the overall variance and that it is associated with a Eigen value greater than 1 (1.91) and with the following factor loadings: parents' employment status (HISEI) 0.85; their level of education (PARED) 0.82; family background (HOMEPOS) 0.73. The ESCS of each student was then calculated by centering the coordinates of the first dimension, so that a student with a strictly positive individual ESCS value is a student with a more favourable socio-economic-cultural background than the average of the study participants. The size of the deviation from the average of the students participating in the study can be assessed as a function of the standard deviation.

### **Metadata regarding the sociolinguistic background of the participating students**

The language background of the students is documented for the following aspects.

#### **Language used in different contexts**

The questionnaire collected detailed information on the language(s) that is/are present in a student's environment and thus used and heard most frequently across different contexts. These are recorded in the items starting with `author_context_language_*` and were elicited with the following prompt:

"Indica, per favore, se il tuo ambiente familiare, scolastico, la tua cerchia di amicizie, ecc. sono/sono stati unicamente o prevalentemente italiani oppure se l'ambiente in cui vivi, vai a scuola, coltivi le tue amicizie è "altro", ovvero arricchito da altre lingue e culture che possono essere quelle "tipiche" altoatesine - tedesca e ladina -, oppure lingue e culture diverse come quella spagnola, inglese, albanese, cinese, araba, eccetera. Per rispondere, pensa soprattutto agli ultimi 10 anni della tua vita (più o meno dall'inizio della scuola primaria) e leggi bene quali sono le possibili risposte."

- `author_context_language_family`: l'ambiente familiare in senso allargato (ovvero includendo parenti e amici frequentati abitualmente dalla tua famiglia) è/è stato
- `author_context_language_school`: Complessivamente, l'ambiente scolastico è/è stato
- `author_context_language_friends`: Complessivamente, il contesto delle amicizie è/è stato
- `author_context_language_free_time`: Complessivamente, il contesto del tempo libero (es. associazioni sportive, parrocchia, centri giovani, associazioni culturali, ecc.), è/è stato

Possible responses were:

- Unicamente italiano
- Più italiano che altro
- Italiano e altro in egual misura
- Più altro che italiano
- Unicamente altro

### Habitual language

The questionnaire collected detailed information on the language(s) that is/are used with different people. These are recorded in the items starting with `author_habitual_language_*` and were elicited with the following prompt:

In quale lingua parli abitualmente con... Scegli l'opzione "non si applica" in caso di assenza della relativa figura familiare.

- `author_habitual_language_parent1`: Genitore 1
- `author_habitual_language_parent2`: Genitore 2
- `author_habitual_language_siblings`: Fratelli/sorelle
- `author_habitual_language_relatives`: Parenti
- `author_habitual_language_friends`: Amici

Possible responses were:

- Unicamente in italiano
- Più italiano che in altra/e lingua/e

- In equal misura in italiano e in altra/e lingua/e
- Più in altra/e lingua/e che in italiano
- Unicamente in altra/e lingua/e
- Non si applica

### Language of education

The questionnaire also collected data on the languages primarily used in educational contexts for different phases of a student's educational path. These are recorded in the metadata items `author_schooling_language_*` and were elicited with the following prompt:

Nella tabella sottostante indica, per favore, quali scuole hai frequentato finora:

<code>author_schooling_language_kindergarten</code>	Scuola dell'infanzia
<code>author_schooling_language_primary</code>	Scuola elementare
<code>author_schooling_language_middle</code>	Scuola media
<code>author_schooling_language_high</code>	Scuola superiore

Possible responses were:

- in lingua italiana
- in lingua tedesca
- in lingua ladina
- in altra lingua (ad es. in un altro Paese)

### First language (L1)

Students have been asked to self-declare their first language, and the first language of their parents, based on which language(s) they know best. Metadata regarding the student's and parents' first language are recorded in the items: `author_L1`, `parent1_L1` and `parent2_L1`. This information has been elicited with the following prompt:

Nella tabella sottostante indica, per favore, quale consideri essere la tua prima lingua, ovvero la lingua che conosci meglio. Fai lo stesso pensando ai tuoi genitori. Scegli l'opzione "non si applica" in caso di assenza della figura genitoriale.

Se hai indicato "Altra lingua" o "Due o più lingue", ti preghiamo di specificare quale/i:

Possible responses included: Italiano | Tedesco | Ladino | Altra lingua | Due o più lingue

Since this question prompted a multitude of different language combinations, the final L1 labels have been simplified recording only the official languages of South Tyrol, Italian (ITA), German (DEU) and Ladin (LLD), and binning all other languages with a placeholder OTHER to maintain student's anonymity. Students indicating more than one language as their or their parents' first language are marked with a + indicating which languages they combine. Occasionally information could not be reconciled leading to missing values.

## Language environment

Moreover, a number of questions aimed at recording the language environment a student was surrounded with, asking for their residence and their or their parents' birthplace as well as when they arrived in South Tyrol in case of people being born abroad or in another Italian region.

<b>author_residence</b>	In quale comune altoatesino vivi attualmente? (Se NON abiti in Alto Adige, seleziona la voce "Non si applica")
<b>author_birthplace</b> <b>author_birthplace_abroad</b>	Dove sei nato/a? Se sei nato/a all'estero, ti preghiamo di specificare il Paese.
<b>author_age_arrival_southtyrol</b>	Se non sei nato/a in provincia di Bolzano, a che età sei arrivato/a in Alto Adige?
<b>parent1_birthplace</b> <b>parent1_birthplace_abroad</b>	Dove sono nati i tuoi genitori? Seleziona l'opzione "non si applica" in assenza della figura genitoriale. GENITORE1
<b>parent2_birthplace</b> <b>parent2_birthplace_abroad</b>	Dove sono nati i tuoi genitori? Seleziona l'opzione "non si applica" in assenza della figura genitoriale. GENITORE 2

## Language assessment score

Finally, a language assessment test (TVI, Test di Verifica in Ingresso from University of Bergamo) was conducted with participating students. The final score is recorded in the metadata item `author_tvi_score`. Values can range from 0 to 28.

## Metadata on reading and writing habits

During the ITACA project, information on reading and writing habits of study participants has been collected in an online questionnaire and can be found in `rw_habits.csv` (in metadata). The collected items are detailed below.

### Reading attitudes

`author_reading_attitudes_*` items are intended to capture the reading attitudes of students. They were translated from Schutte et al. (2007).

They were administered with the following request:

Le seguenti affermazioni riguardano il tuo rapporto con la lettura. Indica, per favore, il tuo grado di accordo o disaccordo con ciascuna affermazione, cercando di prendere posizione e di limitare il più possibile la risposta 'né d'accordo né in disaccordo'.

The items are recoded as:

<b>author_reading_attitudes_1</b>	Se un libro o un articolo è interessante, non mi importa quanto è difficile da leggere.
<b>author_reading_attitudes_2</b>	Senza la lettura la mia vita non sarebbe la stessa.
<b>author_reading_attitudes_3</b>	I miei amici/le mie amiche a volte sono sorpresi da quanto leggo.
<b>author_reading_attitudes_4</b>	A me e ai miei amici/alle mie amiche piace scambiarsi libri o articoli che ci piacciono particolarmente.

<b>author_reading_attitudes_5</b>	Per me è molto importante passare del tempo a leggere.
<b>author_reading_attitudes_6</b>	Rispetto ad altre attività, per me la lettura è importante.
<b>author_reading_attitudes_7</b>	Se ho bisogno di informazioni dal materiale che leggo, finisco di leggere molto prima di quando devo studiare e imparare ciò che leggo.
<b>author_reading_attitudes_8</b>	I miei voti a scuola rispecchiano quanto è efficace il mio modo di leggere.
<b>author_reading_attitudes_9</b>	Attraverso il mio rapporto con la lettura rappresento un buon esempio per gli altri.
<b>author_reading_attitudes_10</b>	Leggo velocemente.
<b>author_reading_attitudes_11</b>	La lettura mi aiuta a dare un senso alla mia vita.
<b>author_reading_attitudes_12</b>	Per me è importante ricevere complimenti per le conoscenze che acquisisco leggendo.
<b>author_reading_attitudes_13</b>	Mi piace che gli altri mi facciano domande su ciò che leggo, perché così posso mostrare la mia conoscenza.
<b>author_reading_attitudes_14</b>	Non mi piace leggere materiale tecnico (manuali, libretti di istruzioni, ecc.). [REVERSED ITEM]
<b>author_reading_attitudes_15</b>	Per me è importante che gli altri notino quanto leggo.
<b>author_reading_attitudes_16</b>	Mi piacciono i libri o gli articoli difficili e impegnativi.
<b>author_reading_attitudes_17</b>	Non mi piace leggere materiale con un vocabolario difficile. [REVERSED ITEM]
<b>author_reading_attitudes_18</b>	Faccio tutte le letture previste per la scuola.
<b>author_reading_attitudes_19</b>	Sono sicuro di poter capire libri o articoli difficili.
<b>author_reading_attitudes_20</b>	Sono un buon lettore/una buona lettrice.
<b>author_reading_attitudes_21</b>	Leggo per migliorare il mio rendimento a scuola.

The 5-point Likert scale has not been recoded.

## Writing attitudes

`author_writing_attitudes_*` items aim to capture the writing attitudes of students. They were translated from Troia et al. (2013). Original English items are not published with the article. They were shared by the authors with the ITACA corpus compilers.

Items for this category were administered with the following request:

Le affermazioni che seguono riguardano il tuo rapporto con la scrittura. Indica quanto ciascuna delle affermazioni è vera per te.

The items are recoded as:

<b>author_writing_attitudes_1</b>	Mi piace scrivere testi impegnativi che mi fanno riflettere.
<b>author_writing_attitudes_2</b>	Credo di saper scrivere dei buoni temi.
<b>author_writing_attitudes_3</b>	Preferisco più leggere che scrivere. [REVERSED ITEM]
<b>author_writing_attitudes_4</b>	È importante per me che il mio insegnante dica agli altri che so scrivere bene.
<b>author_writing_attitudes_5</b>	Mi piace scrivere, così posso imparare di più sugli argomenti che mi interessano.
<b>author_writing_attitudes_6</b>	Cerco di scrivere il meno possibile. [REVERSED ITEM]

<b>author_writing_attitudes_7</b>	Quando scrivo un tema, mi risulta facile trovare degli argomenti da includere in esso.
<b>author_writing_attitudes_8</b>	Scrivere può essere molto divertente.
<b>author_writing_attitudes_9</b>	Mentre rileggo e modifico quanto ho scritto, trovo difficile individuare tutti gli errori. [REVERSED ITEM]
<b>author_writing_attitudes_10</b>	Scrivo soprattutto perché devo farlo per la scuola. [REVERSED ITEM]
<b>author_writing_attitudes_11</b>	Se prendo un buon voto in un compito di scrittura, è perché ho avuto fortuna. [REVERSED ITEM]
<b>author_writing_attitudes_12</b>	Mi piace scrivere su argomenti nuovi.
<b>author_writing_attitudes_13</b>	Nei temi cerco di ottenere il voto più alto della classe.
<b>author_writing_attitudes_14</b>	Se un compito di scrittura è interessante, non mi importa quanto è difficile.
<b>author_writing_attitudes_15</b>	Mi piace ricevere complimenti per come scrivo.
<b>author_writing_attitudes_16</b>	Quando in classe ci viene chiesto di scrivere un saggio, una relazione o una storia, il mio testo è uno dei migliori.
<b>author_writing_attitudes_17</b>	So scrivere dei buoni temi perché scrivere per me è facile.
<b>author_writing_attitudes_18</b>	Imparare a scrivere bene mi aiuterà a diventare un adulto di successo.
<b>author_writing_attitudes_19</b>	Quando scrivo un testo, è difficile per me decidere cosa va al primo, secondo, terzo posto eccetera. [REVERSED ITEM]
<b>author_writing_attitudes_20</b>	Se prendo un buon voto in un tema, è perché mi sono impegnato molto. [REVERSED ITEM?]
<b>author_writing_attitudes_21</b>	Quando scrivo un tema, è facile per me scrivere le mie idee usando frasi ben strutturate.
<b>author_writing_attitudes_22</b>	Mi piace scrivere.
<b>author_writing_attitudes_23</b>	Se non devo correggere quello che ho scritto, ne sono felice. [REVERSED ITEM]
<b>author_writing_attitudes_24</b>	Quando scrivo un tema, faccio fatica a trovare le parole giuste per esprimere quello che voglio dire. [REVERSED ITEM]
<b>author_writing_attitudes_25</b>	Mi sento realizzato se vedo che le mie capacità di scrittura sono migliorate molto.
<b>author_writing_attitudes_26</b>	Per me è importante mostrare agli altri che sono in gamba, attraverso ciò che scrivo.
<b>author_writing_attitudes_27</b>	Preferirei non avere compiti scritti (temi, relazioni, ecc.) da fare a casa. [REVERSED ITEM]
<b>author_writing_attitudes_28</b>	Riuscire a esprimere per iscritto un concetto davvero difficile o confuso mi dà molta soddisfazione.
<b>author_writing_attitudes_29</b>	Non mi piace dover svolgere compiti di scrittura difficili. [REVERSED ITEM]
<b>author_writing_attitudes_30</b>	Non prendo spesso buoni voti negli scritti perché non sono abbastanza intelligente. [REVERSED ITEM]

The 5-point Likert scale has not been recoded.

## Writing habits

author\_writing\_habits\_\* items intend to capture the types of writing most written by students. The items were taken from questionnaires used by the Institute for Applied Linguistics

for the project “KoKo: Eine korpusunterstützte Analyse der Sprachkompetenzen bei Lernenden im deutschen Sprachraum unter besonderer Berücksichtigung des Deutschen in Südtirol”.

They were administered with the following request:

In una settimana, mediamente con quale frequenza scrivi i seguenti testi? Una risposta per singola voce.

The response items are recoded as:

<b>author_writing_habits_email</b>	E-mail
<b>author_writing_habits_social</b>	Post sui Social Network (Facebook, Instagram, Tumblr, ecc.)
<b>author_writing_habits_blog</b>	Blog post
<b>author_writing_habits_writers_forum</b>	Testi su siti/social network per scrittori (Wattpad, The Incipit, Penne Matte, ecc.)
<b>author_writing_habits_whatsapp</b>	Messaggi istantanei (WhatsApp, Telegram, Messenger, Viber, ecc.)
<b>author_writing_habits_twitter</b>	Tweet (Twitter)
<b>author_writing_habits_diary</b>	Pagine di diario personale
<b>author_writing_habits_literature</b>	Poesie, racconti
<b>author_writing_habits_songs</b>	Canzoni
<b>author_writing_habits_other</b>	*Se ti dedichi alla scrittura di altri generi testuali, ti preghiamo di specificare a che genere di testo fai riferimento:

The frequency scale (5-points, from never to always) has not been recoded.

The single item `author_writing_publish` was meant to ask whether the student ever published one of his writings either online or in print.

It was administered with the following request:

Hai mai scritto un testo (ad es. racconto, poesia, saggio, romanzo, ecc.) da pubblicare a stampa oppure online su siti ad hoc o su un blog personale?

## Reading habits

`author_reading_habits_*` items and `parents_reading_habits_*` items intend to capture the reading habits of the students and their parents: Whether students read outside school, which kind of books, online or printed, and whether they read newspapers, magazines, etc.

All items are taken from "Indagine multiscopo sulle famiglie - I cittadini e il tempo libero 2015" by ISTAT (Istituto Nazionale di Statistica) and were adapted by Chiara Vettori to fit the target population.

### Outside school (extracurricular)

The `author_reading_habits_no_school` is a yes/no question that asks whether a student reads outside the school. To this question a logic is applied.

- YES: the student continues answering to the `author_reading_habits_no_school_*` items.
- NO: the student is required to give a reason why they does not read outside the school, continuing to the `author_reading_habits_no_school_why_*`

The `author_reading_habits_no_school` was administered with the following request:

"Negli ultimi 12 mesi hai letto dei libri, in formato cartaceo o elettronico o ascoltato degli audiolibri? Considera solo i libri letti per motivi non strettamente scolastici (= escludi i romanzi e/o le e opere che hai letto per "obbligo scolastico")."

If NO, provide reasons, prompted with the following request:

"Puoi indicare i motivi per cui negli ultimi 12 mesi non hai letto libri per motivi non strettamente scolastici?"

The response items are recoded as columns with these names. Values in the columns are the same as the original response items, except for the last one, i.e. 'other (please specify)', for which open answers have been collected from students.

<code>author_reading_habits_no_school_why_1</code>	I libri costano troppo
<code>author_reading_habits_no_school_why_2</code>	Non ho un posto tranquillo dove leggere
<code>author_reading_habits_no_school_why_3</code>	Ho poco tempo libero
<code>author_reading_habits_no_school_why_4</code>	I libri sono scritti in modo difficile
<code>author_reading_habits_no_school_why_5</code>	Sono troppo stanco/a dopo avere studiato/fatto i compiti
<code>author_reading_habits_no_school_why_6</code>	Mi annoia, non mi appassiona
<code>author_reading_habits_no_school_why_7</code>	Preferisco altri svaghi
<code>author_reading_habits_no_school_why_8</code>	Al giorno d'oggi non serve più leggere
<code>author_reading_habits_no_school_why_9</code>	Preferisco altre forme di comunicazione (televisione, radio, computer, cinema)
<code>author_reading_habits_no_school_why_10</code>	Ci vuole troppo tempo, ho bisogno di stimoli più veloci
<code>author_reading_habits_no_school_why_11</code>	È sufficiente essere informati (attraverso giornali, settimanali, riviste)
<code>author_reading_habits_no_school_why_12</code>	Leggo già molti libri per obbligo scolastico
<code>author_reading_habits_no_school_why_13</code>	Altro (ti preghiamo di specificare)

## Languages

If YES, they are asked whether they also read in other

languages: `author_reading_habits_no_school_languages`, with the following question:

Hai letto o ascoltato anche libri in una lingua diversa dall'italiano?

`author_reading_habits_no_school_languages` is a yes/no question.

If yes, they are asked which

languages: `author_reading_habits_no_school_which_languages`, with the following request:

Se sì, in quale lingua? Sono possibili più risposte.

And with the following items recodings:

<b>author_reading_habits_no_school_which_languages_de</b>	Tedesco
<b>author_reading_habits_no_school_which_languages_lad</b>	Ladino
<b>author_reading_habits_no_school_which_languages_en</b>	Inglese
<b>author_reading_habits_no_school_which_languages_fr</b>	Francese
<b>author_reading_habits_no_school_which_languages_sp</b>	Spagnolo
<b>author_reading_habits_no_school_which_languages_al</b>	Albanese
<b>author_reading_habits_no_school_which_languages_ar</b>	Arabo
<b>author_reading_habits_no_school_which_languages_ru</b>	Russo
<b>author_reading_habits_no_school_which_languages_other</b>	Altro (ti preghiamo di specificare)

### Formats

If NO, they go directly to `author_reading_formats_no_school_*` which is a question that asks whether online formats (e-books, online books and audiobooks) have been read in the past 12 months.

Se pensi al solo formato elettronico/digitale, negli ultimi 12 mesi ti è capitato di fare uso dei seguenti formati di lettura? Nel rispondere, considera sempre solo i libri a cui ti sei avvicinato/a per piacere/interesse personale (= escludi i romanzi e/o le opere che hai letto per “obbligo scolastico”).

Items are recoded into following columns with yes/no answers:

<b>author_reading_formats_no_school_ebook</b>	E-book
<b>author_reading_formats_no_school_onlinebooks</b>	Libri online
<b>author_reading_formats_no_school_audiobooks</b>	Audiolibri

The same options are given to answer `author_reading_or_listening_no_school`, which is prompted by:

In generale preferisci leggere o ascoltare?

### Genres

#### *Literary and technical prose*

`author_reading_habits_no_school_genre_*` items intend to capture the genres (high culture) most read by students.

They were administered with the following request:

Quali dei seguenti generi di libri ti è capitato di leggere o ascoltare, per motivi non strettamente scolastici, negli ultimi 12 mesi? Sono possibili più risposte.

The response items are recoded in separate columns that correspond to:

<b>author_reading_habits_no_school_genre_novel</b>	Romanzi, racconti, poesia, teatro
<b>author_reading_habits_no_school_genre_romance</b>	Romanzi rosa
<b>author_reading_habits_no_school_genre_thriller</b>	Gialli, noir

<b>author_reading_habits_no_school_genre_scifi</b>	Fantascienza
<b>author_reading_habits_no_school_genre_fantasy</b>	Fantasy, horror
<b>author_reading_habits_no_school_genre_socialsci</b>	Libri di scienze sociali o umane (filosofia, sociologia, politica, psicologia, storia, pedagogia, ecc.)
<b>author_reading_habits_no_school_genre_natsci</b>	Libri di scienze naturali, esatte, applicate, di tecnica
<b>author_reading_habits_no_school_genre_art</b>	Arte
<b>author_reading_habits_no_school_genre_religion</b>	Religione
<b>author_reading_habits_no_school_genre_music</b>	Musica
<b>author_reading_habits_no_school_genre_instant</b>	Libri di attualità (instant book)
<b>author_reading_habits_no_school_genre_other:</b>	Altro (ti preghiamo di specificare)

### *Popular genres*

`author_reading_habits_no_school_genre_other_*` items intend to capture the genres (pop culture) most read by students.

They were elicited with the following request:

Oltre a quelli già elencati, di quali altri generi letterari hai letto o ascoltato delle opere, per motivi non strettamente scolastici, negli ultimi 12 mesi? Sono possibili più risposte.

The response items are recoded in separate columns that correspond to:

<b>author_reading_habits_no_school_genre_other_humor</b>	Umoristici
<b>author_reading_habits_no_school_genre_other_hobby</b>	Hobby e tempo libero
<b>author_reading_habits_no_school_genre_other_astro</b>	Astrologia, magia, esoterismo
<b>author_reading_habits_no_school_genre_other_tech</b>	Libri di informatica
<b>author_reading_habits_no_school_genre_other_tourism</b>	Guide turistiche
<b>author_reading_habits_no_school_genre_other_health</b>	Libri sulla salute
<b>author_reading_habits_no_school_genre_other_home</b>	Libri per la casa (cucina, bricolage, maglia, cucito, ecc.)
<b>author_reading_habits_no_school_genre_other_manual</b>	Manuali pratici (ad es. la collana 'For Dummies' [per principianti])
<b>author_reading_habits_no_school_genre_other_photo</b>	Fotografia, cinema
<b>author_reading_habits_no_school_genre_other_comics</b>	Libri a fumetti, graphic novel
<b>author_reading_habits_no_school_genre_other_animals</b>	Libri sugli animali
<b>author_reading_habits_no_school_genre_other_illustr</b>	Albi illustrati
<b>author_reading_habits_no_school_genre_other_bio</b>	Biografie
<b>author_reading_habits_no_school_genre_other_none</b>	Nessun altro genere letterario oltre a quelli già indicati
<b>author_reading_habits_no_school_genre_other_other</b>	Altro (ti preghiamo di specificare)

### *Frequency, quantity and purchase*

The following items give general information on the extracurricular reading habits of students:

author\_reading\_habits\_no\_school\_frequency prompted with:

Con quale frequenza leggi o ascolti libri per motivi non strettamente scolastici?

author\_reading\_habits\_no\_school\_quantity prompted with:

Complessivamente, quanti libri hai letto o ascoltato negli ultimi 12 mesi per motivi non strettamente scolastici?

author\_reading\_habits\_no\_school\_purchase prompted with:

Come sei venuto/a in possesso dell'ultimo libro che hai letto o ascoltato?

Responses have not been recoded.

### At school (curricular)

The author\_reading\_habits\_school is a yes/no question that asks whether a student reads for the curriculum. This question is formulated as:

Considera gli ultimi 12 mesi: hai letto o ascoltato libri per motivi scolastici (esclusi i libri di testo obbligatori)?

To this question a logic is applied.

- YES: the student answers the author\_reading\_habits\_school\_quantity item. This question is formulated as:

Quanti libri hai letto o ascoltato?

Answers are not recoded.

- NO: the student goes directly to the newspaper section author\_reading\_habits\_newspaper\_\*

### Newspapers

The author\_reading\_habits\_newspaper\_\* items are about newspaper reading habits by the students.

The author\_reading\_habits\_newspaper item asks whether and how often students read newspapers. It is prompted as:

Leggi quotidiani cartacei o online almeno una volta alla settimana?

A logic is applied:

- If not "NO", then students answer to the author\_reading\_habits\_newspaper\_genre\_\*items

Che tipo di articoli leggi prevalentemente sui quotidiani?

The response items are recoded in separate columns that correspond to:

author_reading_habits_newspaper_genre_natpol	Politica nazionale
author_reading_habits_newspaper_genre_internatpol	Politica internazionale

<b>author_reading_habits_newspaper_genre_economy</b>	Economia, finanza
<b>author_reading_habits_newspaper_genre_report</b>	Cronaca
<b>author_reading_habits_newspaper_genre_culture</b>	Culturali
<b>author_reading_habits_newspaper_genre_showbiz</b>	Spettacoli
<b>author_reading_habits_newspaper_genre_localnews</b>	Notizie locali
<b>author_reading_habits_newspaper_genre_latnews</b>	Attualità
<b>author_reading_habits_newspaper_genre_sport</b>	Sport
<b>author_reading_habits_newspaper_genre_tech</b>	Tecnologia, scienze, ambiente
<b>author_reading_habits_newspaper_genre_comments</b>	Approfondimenti sulla cronaca
<b>author_reading_habits_newspaper_genre_editorials</b>	Editoriali
<b>author_reading_habits_newspaper_genre_horoscope</b>	Oroscopo
<b>author_reading_habits_newspaper_genre_post</b>	Rubriche di posta
<b>author_reading_habits_newspaper_genre_other</b>	Altro (ti preghiamo di specificare)

- If no, they go directly to the questions about library services  
(author\_reading\_habits\_library\_\*)

### Library services

The author\_reading\_habits\_library\_\* items are about use of library services by the students.

The author\_reading\_habits\_library item asks whether students have been to or used library services in the last year. It is prompted as:

Negli ultimi 12 mesi sei stato/a in una biblioteca o hai usufruito dei servizi di una biblioteca?

A logic is applied:

- If YES, then students answer to  
the author\_reading\_habits\_library\_services\_\* items, prompted as:

Per quali delle seguenti attività sei stato/a in biblioteca o hai usufruito dei servizi di una biblioteca? (Sono possibili più risposte)

The response items are recoded in separate columns that correspond to:

<b>author_reading_habits_library_services_catalogues</b>	Per consultare cataloghi
<b>author_reading_habits_library_services_booking</b>	Per prenotare prestiti
<b>author_reading_habits_library_services_loan</b>	Per prendere in prestito libri
<b>author_reading_habits_library_services_audiovisual</b>	Per prendere in prestito materiale audio-visivo (CD-rom, DVD, ecc.)
<b>author_reading_habits_library_services_mlol</b>	Per scaricare e-book, quotidiani, riviste, ecc. (ad es. attraverso la rete di digital lending MLOL)
<b>author_reading_habits_library_services_other</b>	Altro (ti preghiamo di specificare)

- If NO, they fill out the sociolinguistic questionnaire and then answer to the reading habits related to their parents (parents\_reading\_habits\_\* items).

## Reading habits - parents

The `parents_reading_habits_*` items are about the reading habits of the parents.

The `parents_reading_habits` item asks whether the parents of the student read books. It is prompted as:

I tuoi genitori leggono libri (cartacei o elettronici) o ascoltano audiolibri?

A logic is applied:

- If YES, then students answer the `parents_reading_habits_genre_*` items, prompted as:

Che genere di libri leggono o ascoltano i tuoi genitori? Nel rispondere riferisciti al genitore che legge di più fra i due.

The response items are recoded in separate columns that correspond to:

<code>parents_reading_habits_genre_novel</code>	Romanzi, racconti, poesia, teatro
<code>parents_reading_habits_genre_romance</code>	Romanzi rosa
<code>parents_reading_habits_genre_thriller</code>	Gialli, noir
<code>parents_reading_habits_genre_scifi</code>	Fantascienza
<code>parents_reading_habits_genre_fantasy</code>	Fantasy, horror
<code>parents_reading_habits_genre_socialsci</code>	Libri di scienze sociali o umane (filosofia, sociologia, politica, psicologia, storia, pedagogia, ecc.)
<code>parents_reading_habits_genre_natsci</code>	Libri di scienze naturali, esatte, applicate, di tecnica
<code>parents_reading_habits_genre_art</code>	Arte
<code>parents_reading_habits_genre_religion</code>	Religione
<code>parents_reading_habits_genre_music</code>	Musica
<code>parents_reading_habits_genre_instant</code>	Libri di attualità (instant book)
<code>parents_reading_habits_genre_humor</code>	Umoristici
<code>parents_reading_habits_genre_hobby</code>	Hobby e tempo libero
<code>parents_reading_habits_genre_astro</code>	Astrologia, magia, esoterismo
<code>parents_reading_habits_genre_tech</code>	Libri di informatica
<code>parents_reading_habits_genre_tourism</code>	Guide turistiche
<code>parents_reading_habits_genre_health</code>	Libri sulla salute
<code>parents_reading_habits_genre_home</code>	Libri per la casa (cucina, bricolage, maglia, cucito, ecc.)
<code>parents_reading_habits_genre_manual</code>	Manuali pratici (ad es. la collana 'For Dummies' [per principianti])
<code>parents_reading_habits_genre_photo</code>	Fotografia, cinema
<code>parents_reading_habits_genre_comics</code>	Libri a fumetti, graphic novel
<code>parents_reading_habits_genre_animals</code>	Libri sugli animali
<code>parents_reading_habits_genre_illustr</code>	Albi illustrati
<code>parents_reading_habits_genre_bio</code>	Biografie
<code>parents_reading_habits_genre_other</code>	Altro (ti preghiamo di specificare)

- If NO, they directly end the survey.

## Coherence ratings

Manual coherence ratings were obtained for the texts in the ITACA corpus and recorded in `ratings.csv` (see metadata). In the following we describe the individual items of the coherence rating.

Note that `ratings.csv` can contain various ratings per text. To calculate inter-rater agreement and ensure rating quality, most texts received two ratings. However, 12 of the texts were rated by only one rater, because one of the raters decided that those texts were too short to be rated. Ten of the texts were rated in a pilot by 6 raters. Finally, one of the texts (CA31TO\_LAR14V) has 4 ratings, because it was one of three (next to ST23TO\_PER21M and GR25ME\_NAR22A) that has been used for the training of the raters.

Questions were administered with the following request:

Indica, per favore, il tuo grado di accordo o disaccordo con ciascuna affermazione.

The items to rate are recoded as:

<b>likert_001</b>	Il testo è di tipo narrativo o espositivo. [REVERSED ITEM]
<b>likert_002</b>	Il testo rispecchia il genere testuale previsto dalla consegna (lettera).
<b>likert_003</b>	Il registro è generalmente adeguato al tipo di task assegnato (lettera formale a un rappresentante delle istituzioni).
<b>likert_004</b>	La trattazione del tema è aderente al task assegnato (ovvero non c'è confusione tra DAD e DDI).
<b>likert_007</b>	Gli incisi sono di norma correttamente segnalati.
<b>likert_008</b>	Nel testo sono presenti molte virgole splice. [REVERSED ITEM]
<b>likert_005</b>	All'interno del paragrafo la punteggiatura è scarsamente segnalata. [REVERSED ITEM]
<b>likert_006</b>	L'uso del punto nel testo è sempre funzionale alla suddivisione degli enunciati.
<b>likert_009</b>	Il testo è suddiviso in modo incoerente o non è affatto suddiviso in sezioni. [REVERSED ITEM]
<b>likert_010</b>	Il testo è suddiviso in paragrafi coerenti con i cambi di argomento (movimenti testuali).
<b>likert_011</b>	La segmentazione è funzionale all'articolazione del testo.
<b>likert_012</b>	Le devianze nella punteggiatura sono tali da compromettere la coerenza e/o la comprensione del testo. [REVERSED ITEM]
<b>likert_013</b>	Nel testo vengono rispettate le regole di interpunzione.
<b>likert_014</b>	Di norma i connettivi sono impiegati coerentemente con la loro semantica.
<b>likert_015</b>	I connettivi sono spesso impiegati in modo sintatticamente improprio. [REVERSED ITEM]
<b>likert_016</b>	Le devianze nell'uso dei connettivi sono rilevanti. [REVERSED ITEM]
<b>likert_017</b>	Il collegamento tra la ripresa e il suo referente è sempre ricostruibile.

<b>likert_018</b>	I riferimenti (incapsulatori) a contenuti già menzionati nel testo a volte sono difficili da ricostruire. [REVERSED ITEM]
<b>likert_019</b>	Le parti del testo sono per lo più giustapposte. [REVERSED ITEM]
<b>likert_020</b>	L'uso dei connettivi e delle anafore è funzionale alla coesione del testo.
<b>likert_021</b>	Nel testo è riconoscibile un'idea di fondo.
<b>likert_022</b>	Il testo è strutturato sulla base dell'idea di fondo.
<b>likert_023</b>	La tesi è riconoscibile.
<b>likert_024</b>	La tesi è sostenuta da argomenti.
<b>likert_025</b>	Non tutte le argomentazioni sono sviluppate. [REVERSED ITEM]
<b>likert_026</b>	Le conclusioni derivano dagli argomenti presenti nel testo.
<b>likert_027</b>	Le idee presenti nel testo sono ben collegate tra loro.
<b>likert_028</b>	Il testo è costruito a elenco, ovvero è privo di una gerarchia riconoscibile. [REVERSED ITEM]
<b>likert_029</b>	Sono presenti delle devianze nella sequenza logica, tematica o temporale degli argomenti. [REVERSED ITEM]
<b>likert_030</b>	Gli argomenti introdotti nel testo sono ridondanti. [REVERSED ITEM]
<b>likert_031</b>	Il testo è stato progettato in modo efficace.
<b>likert_032</b>	Il contenuto è contraddittorio/incoerente. [REVERSED ITEM]
<b>likert_033</b>	Il contenuto è convincente.
<b>likert_034</b>	Gli argomenti introdotti nel testo sono funzionali allo scopo.
<b>likert_035</b>	Gli argomenti introdotti nel testo vanno fuori tema (off-topic). [REVERSED ITEM]
<b>likert_036</b>	Gli argomenti introdotti nel testo sono generici. [REVERSED ITEM]
<b>likert_037</b>	I termini utilizzati provocano fraintendimenti. [REVERSED ITEM]
<b>likert_038</b>	I termini utilizzati sono coerenti con la loro semantica e il contesto in cui appaiono.
<b>likert_039</b>	Nel testo ci sono informazioni implicite che rendono difficoltosa la comprensione. [REVERSED ITEM]
<b>likert_040</b>	Tutte le informazioni necessarie alla piena comprensione del testo sono ben esplicitate.
<b>likert_041</b>	Il lettore deve sforzarsi per comprendere quanto scritto. [REVERSED ITEM]
<b>likert_042</b>	Sono necessarie diverse riletture per comprendere il testo o alcune sue parti. [REVERSED ITEM]
<b>likert_043</b>	Il testo si comprende facilmente.
<b>likert_044</b>	Lo scopo/gli scopi del testo è/sono chiaro/i.

Ratings were done on a Likert scale recoded as:

"3": "Del tutto d'accordo"

"2": "Un po' d'accordo"

"1": "Un po' in disaccordo"

"0": "Del tutto in disaccordo"

The rating scale for reversed items is recoded as:

"0": "Del tutto d'accordo"

"1": "Un po' d'accordo"

"2": "Un po' in disaccordo"

"3": "Del tutto in disaccordo"

## References

International Labour Organization. (n.d.). International Standard Classification of Occupations (ISCO). *ILOSTAT*. Retrieved June 10, 2025, from <https://ilostat.ilo.org/methods/concepts-and-definitions/classification-occupation/>

Josse J., Husson F. (2016). “missMDA: A Package for Handling Missing Values in Multivariate Data Analysis.” *Journal of Statistical Software*, 70(1), 1–31. doi:10.18637/jss.v070.i01.

Masters, G.N., Wright, B.D. (1997). The Partial Credit Model. In: van der Linden, W.J., Hambleton, R.K. (eds) *Handbook of Modern Item Response Theory*. Springer, New York, NY. [https://doi.org/10.1007/978-1-4757-2691-6\\_6](https://doi.org/10.1007/978-1-4757-2691-6_6)

PISA 2018 Results (Volume III): What School Life Means for Students’ Lives, Annex A1. Construction of indices, <https://www.oecd-ilibrary.org/sites/0a428b07-en/index.html?itemId=/content/component/0a428b07-en#s131>

Schwittler N. (2023). “ISCO08ConveRsions: Converts ISCO-08 to Job Prestige Scores, ISCO-88 and Job Name”, <https://CRAN.R-project.org/package=ISCO08ConveRsions>

Schutte, N. S., & Malouff, J. M. (2007). Dimensions of Reading Motivation: Development of an Adult Reading Motivation Scale. *Reading Psychology*, 28(5), 469–489. <https://doi.org/10.1080/02702710701568991>

Troia, G. A., Harbaugh, A. G., Shankland, R. K., Wolbers, K. A., & Lawrence, A. M. (2013). Relationships between writing motivation, writing activity, and writing performance: effects of grade, sex, and ability. *Reading and Writing*, 26(1), 17–44. <https://doi.org/10.1007/s11145-012-9379-2>

# eurac research

Authors: Bienati, Arianna; Zanasi, Lorenzo

Date: 2024-11-18

Publisher: Eurac Research

Version: 1.1

## Annotation schema for the ITACA corpus

### Introduction

This document describes the procedure followed to manually annotate a subcorpus of 388 texts from the ITACA corpus. This subcorpus has been chosen for annotation as it contains texts written by:

- students from the `sample_repr_south_tyrol` (more info in the [Corpus handbook](#))
- students who attended schools with Italian as the main medium of instruction from 1st to 12th grade
- students who declare Italian as (one of) their L1(s)

The annotated textual traits have been selected through this procedure (more info [here](#)):

1. review of educational linguistics studies from 1990 to 2020 and extraction of features presented as 'textual'. Output: 75 problematic textual features found in students' texts (from primary school to university)
2. systematization of the traits according to the Basel Model (Ferrari et al. 2021).
3. Pilot annotation in NVivo of argumentative essays written during the ITACA pre-test (reduction to 23 traits + internal structuring and aggregation of some traits).
4. Final selection of the traits to be annotated (9 textual traits): further internal structuring of some traits and lumping of categories (e.g., all accuracy problems with the use of connectives, both semantic and syntactic have been subsumed under the category 'problematic' - problematico).

In what follows we report the original annotation guidelines (written in Italian) and the reading protocol followed by the annotators in the annotation process.

# Schema di annotazione

## Tratti grammaticali

### Accordo

**Node annotation:** AccordoProblematico

**Edge type:** AccordoProblematico

**Edge annotation:** AccProblGenere; **Value:** true, false.

**Edge annotation:** AccProblNumero; **Value:** true, false.

**Edge annotation:** AccProblPersona; **Value:** true, false.

Definizione accordo: si ha accordo quando una parola cambia forma a seconda delle altre parole a cui è legata. Per questo tag, ci soffermiamo sull'accordo soggetto-verbo.

Usare questo tag quando l'accordo sintattico fra soggetto e verbo non è rispettato. Vedi anche [Accordo \(prontuario\)](#) di Stefano Telve.

Casistica:

- Mancato accordo di genere (AccProblGenere): es. *la ragione (...) è dovuto a...* (Corpus ITACA).

### Warning

ATTENZIONE: se ci sono dubbi sull'accettabilità di certe forme di accordo del participio passato, fare riferimento allo specchietto riassuntivo disponibile sull'[Accademia della Crusca](#), che riassume un contributo di Serianni sul tema del 1994.

Riportiamo qui per comodità i passaggi più salienti:

Il problema dell'accordo del participio passato è uno dei capitoli più spinosi della sintassi italiana. Le principali incertezze possono essere schematizzate nel modo seguente:

1. accordo del participio d'un verbo composto con l'ausiliare avere col complemento oggetto posposto ("ho scelto le migliori opere" - "ho scelte le migliori opere": nettamente prevalente, e quindi anche preferibile, la prima soluzione);
2. accordo del participio d'un verbo composto con avere con l'oggetto anteposto, costituito da un pronome personale o relativo ("ci ha ingannato" - "ci ha ingannati", "la casa che ho comprato" - "la casa che ho comprata"): accettare entrambe.
3. accordo del participio di essere o di un verbo copulativo col soggetto o col nome del predicato ovvero col complemento predicativo ("il suo discorso è stato, è risultato una sorpresa" - "è stata, è risultata una sorpresa"): accettare entrambe.

4. accordo del participio d'un verbo pronominale col soggetto o col complemento oggetto, sia esso anteposto o posposto ("la meta che ci siamo prefissati" - "che ci siamo prefissata": accettare entrambe.

- Mancato accordo di numero: *la produttività, la formazione, l'educazione (...) ci era stata tolta* (Corpus ITACA).

### Warning

ATTENZIONE: In italiano, l'accordo semantico (o anche concordanza a senso) è possibile (v. neostandard), ma abbiamo deciso di segnalarlo comunque: "di fatto sono il 12% delle famiglie di studenti a dichiarare che ..." (Corpus ITACA).

- Mancato accordo di persona: *sia noi adolescenti che i bambini hanno riscontrato* (Corpus ITACA).

## Tratti testuali

### Segmentazione

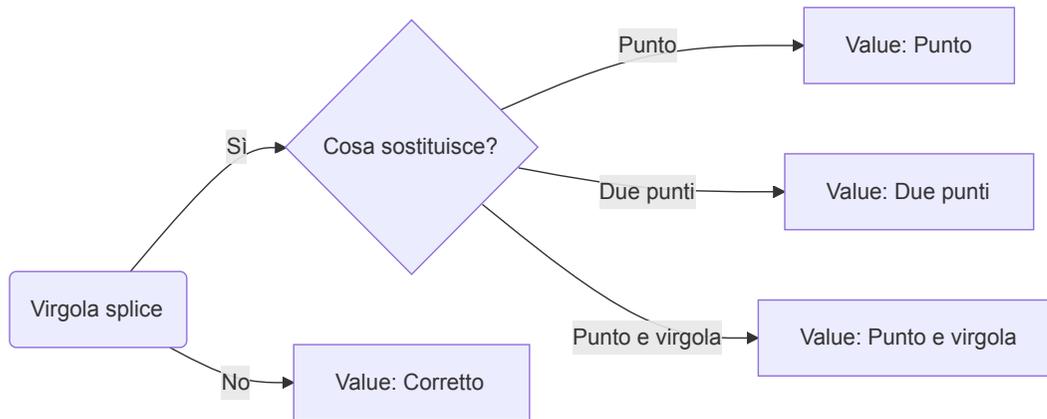
#### Virgola splice

**Node annotation:** VirgolaSplice; **Values:** Corretto, Punto, Due punti, Punto e virgola

Definizione "virgola splice": uso di una virgola per unire due frasi indipendenti. La virgola può apparire in sostituzione a punto, due punti e punto e virgola (Roggia 2010; Colombo 2011; INVALSI 2012; Cignetti 2015; Prada 2016).

L'anglicismo virgola splice (comma splice) individua un fenomeno presente nell'italiano contemporaneo: un particolare collegamento fra parti di testo dovuto all'iper-estensione dell'uso della virgola come separatore in luogo sia di segni di interpunzione di livello superiore (punto, punto e virgola, ma anche due punti), sia di appropriati connettivi.

Procedura di annotazione:



1. Determinare se la virgola è una virgola splice. Domanda a cui devo rispondere: le frasi collegate dalla virgola possono essere lette come frasi indipendenti?

1. Se no, lasciare su "Corretto"
2. Se sì, quale valore sostituisce?

- due punti (:) si inseriscono quando "il secondo enunciato è funzionalizzato al primo ed è necessario per l'interpretazione globale della sequenza" (Ferrari et al. 2018: 156). Le relazioni che possono sussistere tra i due enunciati e che chiamano la normalizzazione con i due punti al posto della splice sono principalmente la specificazione e la motivazione. Per approfondimenti, si veda Ferrari et al. (2018: 156ss).

*È questo che mi piace dello sport, (: due punti) quando ci sei dentro è difficile venirne fuori* (Roggia 2010)

- punto/punto e virgola (./); si inseriscono quando la splice ha la funzione di segmentare il testo in enunciati. Il punto e virgola, in particolare, si può sostituire alla splice quando non marca enumerazioni e articola "due enunciati che intrattengono un legame semantico unitario con il contesto, e che stanno tipicamente in rapporto di opposizione". vedi Ferrari et al. (2018: 69ss, 83ss):

*In questi versi, Ettore, dimostra tutto il suo amore nei confronti del figlio, (. punto) l'uomo dal cuore impavido, si toglie l'elmo e si "sottomette" da un certo punto di vista, (. punto) innalza il figlio tra le sue braccia e lo affida agli dei, (. punto) questa a mio avviso è la cosa più bella che possa fare un padre al proprio figlio, desiderare che divenga migliore di lui stesso.* (Roggia 2010: 204).

## Virgola di troppo

**Node annotation:** VirgolaDiTroppo; **Values:** true, false

Usare l'attributo VirgolaDiTroppo = "true" (= da eliminare) se c'è una virgola nei seguenti contesti:

- Virgola tra soggetto (o completiva soggettiva) e predicato (maggior parte delle occorrenze) (De Masi, Maggio 2008; Roggia 2010; Colombo 2011; INVALSI 2012):

*In questi versi, Ettore, (x eliminare) dimostra (...) l'uomo dal cuore impavido, (x eliminare) si toglie l'elmo ...* (Roggia 2010).

- Virgola tra argomenti nucleari del verbo o del sintagma nominale (es. complementi di specificazione) (Colombo 2011):

*Spesso coloro che spendono maggior tempo davanti alla televisione tendono a far coincidere il mondo rappresentato al suo interno, (x eliminare) con il mondo reale.* (Colombo 2011)

*per un terzo dei bambini e degli adolescenti, (x eliminare) di età compresa tra i 6 e i 17 anni* (Corpus ITACA)

- Virgola che segnala una relativa restrittiva (De Masi, Maggio 2008; Colombo 2011; INVALSI 2012)

*il passeggero, (x eliminare) che nell'ode "La caduta" accorre ad aiutarlo, (x eliminare) cerca di ...* (Colombo 2011)

- Virgola tra due frasi/sintagmi nominali coordinati dalla e:

*Il Sussidiario.net espone le sue riflessioni sull'utilizzo prolungato delle apparecchiature elettroniche può portare alla perdita irreversibile dei neuroni, (x eliminare) e possono comparire delle psicopatologie sociali.* (Corpus ITACA)

## Incisi

**Node annotation:** Inciso; **Values:** Correttamente segnalato, Non segnalato, Manca virgola iniziale, Manca virgola finale

Definizione inciso non segnalato: l'assenza della virgola (o della parentesi o del trattino) in apertura e/o chiusura degli incisi.

Definizione di inciso: breve frase, membretto di periodo che viene inserito, con funzione parentetica, in un costrutto da cui è sintatticamente indipendente.

L'inciso può isolare:

- sintagmi o frasi relative appositive  
*Cosa poteva saperne Tommaso, **che da poco aveva compiuto un anno**, delle nuvole nere che opprimevano il cuore di sua madre?* (Ferrari et al. 2018)
- sintagmi o subordinate circostanziali  
*Ma Lorenzin, **nonostante le polemiche e gli errori**, piace molto ai medici che non indossano la casacca politica sopra il camice)* (Ferrari et al. 2018)

Gli incisi nei testi di ITACA possono essere formalmente segnalati da:

1. due virgole (Value: Correttamente segnalato)
2. oppure possono essere o parzialmente o del tutto NON formalmente segnalati (De Masi, Maggio 2008; Colombo 2011; INVALSI 2012; De Santis, Gatta 2013):
  - una virgola in apertura, ma non in chiusura (Value: Manca virgola finale).

Comprendo la comodità delle lezioni a distanza, **in modo da stimolare multitasking nei giovani, un maggiore controllo dello studio e delle attività extrascolastiche** ma rimango convinta che ... (Corpus ITACA).

- una virgola in chiusura, ma non in apertura (Value: Manca virgola iniziale).

io come moltissimi altri studenti **dato il contesto in cui ci si trova, ovvero casa propria**, tendiamo a non prestare tanta attenzione quando i professori tengono la lezione. (Corpus ITACA).

- no virgole (Value: Non segnalato).

Il mio vicino di casa **che è molto salutista** esige che le persone che fumano **quando vanno a trovarlo** non portino con sé le sigarette.

Lo trovo inammissibile, (x eliminare) che **superato il periodo di pandemia** possa essere presa la decisione di mantenere uno stile di vita, (x eliminare) adottato esclusivamente per tempi che non consentivano altra soluzione. (Corpus ITACA).

#### Posizioni più frequenti degli incisi:

- tra soggetto (nominale o frasale) e predicato,
- tra argomenti del verbo e del nome (sia che siano nominali, sia che siano frasali: es. completeive oggettive)

#### Casi da NON annotare:

- **Inciso monorematico** (composto da una congiunzione o da un avverbio) posto all'inizio o nel mezzo di una frase. Questo inciso può essere o meno seguito da virgola (se in posizione iniziale) o compreso fra virgole (se in posizione mediana).

**Infatti**, molte volte si è parlato di riduzione della capacità degli studenti di seguire le lezioni davanti al computer e, **sicuramente**, questo non dimostra quanto possa essere efficace la DAD.

#### Warning

ATTENZIONE: Fare un'eccezione per l'inciso monorematico semanticamente "pieno". In questi casi, valutare in base alla semantica: se la sequenza si può togliere perché la semantica non cambia, va segnalato come inciso. Sui casi dubbi, la decisione è di chi annota (con attenzione alla semantica e all'impatto sulla coerenza): la mancanza di dispositivi formali (es. virgole, parentesi, trattini) impatta sulla comprensibilità della frase?

Es. "Come ultima problematica vorrei semplicemente rammentare il problema della fusione scuola-vita privata, di ambiti che a mio parere dovrebbero rimanere ben distinti per evitare che si rinunci sempre più al tempo libero essendo legati da uno spazio lavorativo e (x virgola?) viceversa (x virgola?) per evitare che le tentazioni della vita privata influiscano sulle prestazioni scolastiche." (Corpus ITACA).

- **Appendice finale**, in cui la seconda virgola è assorbita da un segno di interpunzione superiore.

*Del Corriere dei Piccoli mi ricordo assolutamente tutto, anche perché avendo il culto della memoria ne ricompro sempre su bancarelle.*

### Warning

ATTENZIONE ALLE VIRGOLE SPLICE! Capita che l'appendice finale non sia segnalata da un segno di interpunzione superiore ma da una virgola splice. Prima normalizzare la splice con il segno di punteggiatura più adatto.

Es. "A mio avviso questo argomento non regge, o almeno non è molto rilevante, (: due punti) le ore scolastiche non cambiano quindi il tempo libero è lo stesso medesimo di quello in presenza." (Corpus ITACA).

- **Quadro** all'inizio della frase (= segmento che precede il nucleo, con funzione di cornice per l'interpretazione denotativa, modale o illocutiva del Nucleo e per cogliere l'aggancio testuale dell'Enunciato con il contesto precedente o successivo)

*Durante questi due anni di pandemia **da quando è scoppiata l'emergenza sanitaria**, tutti gli studenti italiani e non hanno passato del periodo scolastico in DaD provando una nuova esperienza nel proprio percorso scolastico.* (Corpus ITACA).

## Piano logico-argomentativo

### Connettivi

**Node annotation:** ConnProblematico; **Values:** true, false

**Related annotation:** ConnProblComm

Definizione di connettivo: il termine connettivo indica ciascuna delle forme invariabili (congiunzioni, locuzioni, ecc.), che indicano relazioni che strutturano 'logicamente' i significati della frase e del testo (Ferrari 2010; Ferrari 2014; Ferrari et al. 2021).

### Warning

ATTENZIONE: in ITACA sono accolte anche forme che permettono un minimo di variabilità morfologica ("per esempio/ad esempio", "per questo/i motivo/i"). Le forme connettive che sono state pre-annotate nel corpus ITACA ricalcano quelle del [LICO](#) (Lexicon of Italian Connectives).

### Warning

ATTENZIONE: i connettivi marcano relazioni tra eventi. Gli eventi possono essere codificati da verbi, ma anche da nomi (es. dopo il pressante invito = dopo che è stato invitato con insistenza)

Usare l'attributo "true" (= connettivo problematico) quando:

- il connettivo è semanticamente appropriato, ma usato in modo sintatticamente improprio

*Riguardo questo argomento (i.e., la sovrapposizione con altri impegni) posso dire di aver fatto un'esperienza nella mia vita in DAD. (...) **Mentre** con gli orari io mi sono trovata bene, perché appena finite le lezioni potevo già iniziare i compiti e studiare visto che ero già a casa*

Commento: *mentre* è una congiunzione usata come avverbio. Per dubbi sugli usi codificati dei connettivi fare riferimento al [Dizionario di Italiano Sabatini Coletti](#)

- connettivo sintatticamente corretto, ma semanticamente inappropriato (De Masi/Maggio, 2008; Serianni, Benedetti 2009; Colombo, 2011; Ruggiano 2011; Ciccolone 2012; Notarbartolo 2014; Cignetti, 2015; Ruele, Zuin 2020).

*il ragazzo aveva dato forma di squilibrio e (problematico) mai nessuno aveva pensato di ostacolare le sue esercitazioni di volo* (De Masi, Maggio 2008: 169).

Commento: *la e* segnala una relazione concessiva (aspettativa frustrata: se qualcuno dà segni di squilibrio, ci si aspetta che venga fermato prima che compia azioni pericolose).

*Quindi, quello che intendo esporre è che la DAD potrebbe essere un ottimo modo per invogliare gli studenti a crescere personalmente e organizzarsi, **ma** allo stesso tempo per portare il metodo scolastico ad un livello più innovativo e diverso dalla classica spiegazione monotona.*

Commento: *ma* in questo contesto non è usato in senso avversativo, bensì di aggiunta (sostituire con una *e*?).

- costrutti correlativi a cui manca la seconda parte (es. *sia...sia; e...e; da una parte...dall'altra*) (De Masi, Maggio 2008; Invalsi 2012)

## Contraddizione

**Node annotation:** Contraddizione

**Edge type:** Contraddizione\_Relation

**Related annotations:** ContraddizioneComm

Definizione di contraddizione: l'asserzione, entro uno stesso sistema, di una proposizione e la sua negazione (Perelman, Olbrechts-Tyteca 1966:205). In sistemi non formali, come quello linguistico, sarebbe meglio parlare di incompatibilità. C'è incompatibilità quando, date due asserzioni, bisogna sceglierne una delle due, pena la rinuncia ad entrambe. Nei testi scolastici le incompatibilità o contraddizioni sono segnalate in particolare da Serianni, Benedetti (2009), Colombo (2011).

L'annotazione deve segnalare chiaramente i segmenti in contraddizione tra loro. Di solito, si annotano gli span in cui è presente la contraddizione da punto a punto.

A(*Quest'anno a scuola ho conosciuto molti ragazzi e insegnanti nuovi, e sono molto contento perché sono brave persone.*) B(*Solo con due persone della nuova classe ho instaurato un buon rapporto, "Federico" e "Fabio"*) (Serianni, Benedetti 2009: 138).

A(*io non credo di saper precisamente cosa si potrebbe cambiare o come si potrebbe migliorare la situazione*) ... B(*io sono convinta che una soluzione si riesce sempre a trovare, ad esempio mettere un numero massimo di figli per famiglia*) (ITACA pre-test).

### Warning

ATTENZIONE: la contraddizione può essere contenuta anche all'interno di una stessa frase. In questo caso non si annota il fenomeno da punto a punto, ma si annotano le unità informative (di norma delimitate da virgole).

Es. A(in questo caso posso dire che io non mi sentivo così (= più a suo agio a casa),) B(magari meno in ansia lo ero,) C(ma io mi sentivo come in presenza.)

Commento: A e C non sono in contraddizione: In DaD la scrivente non si sentiva più a suo agio, si sentiva come in presenza. B contraddice A e C dicendo che in DaD era meno in ansia. Il 'magari' in B, modalizza la proposizione e sembrerebbe attenuare la gravità della contraddizione.

## Piano tematico-referenziale

### Anafora

**Node annotation:** AnaforaProblematica; **Value:** Ripresa, Antecedente

**Node annotation:** AnTipoRipresa; **Value:** Zero, Pronome, Sintagma nominale

**Node annotation:** AnRipresaScorretta; **Value:** true, false

**Node annotation:** AnTipoAntecedente; **Value:** Ambiguo, Univoco

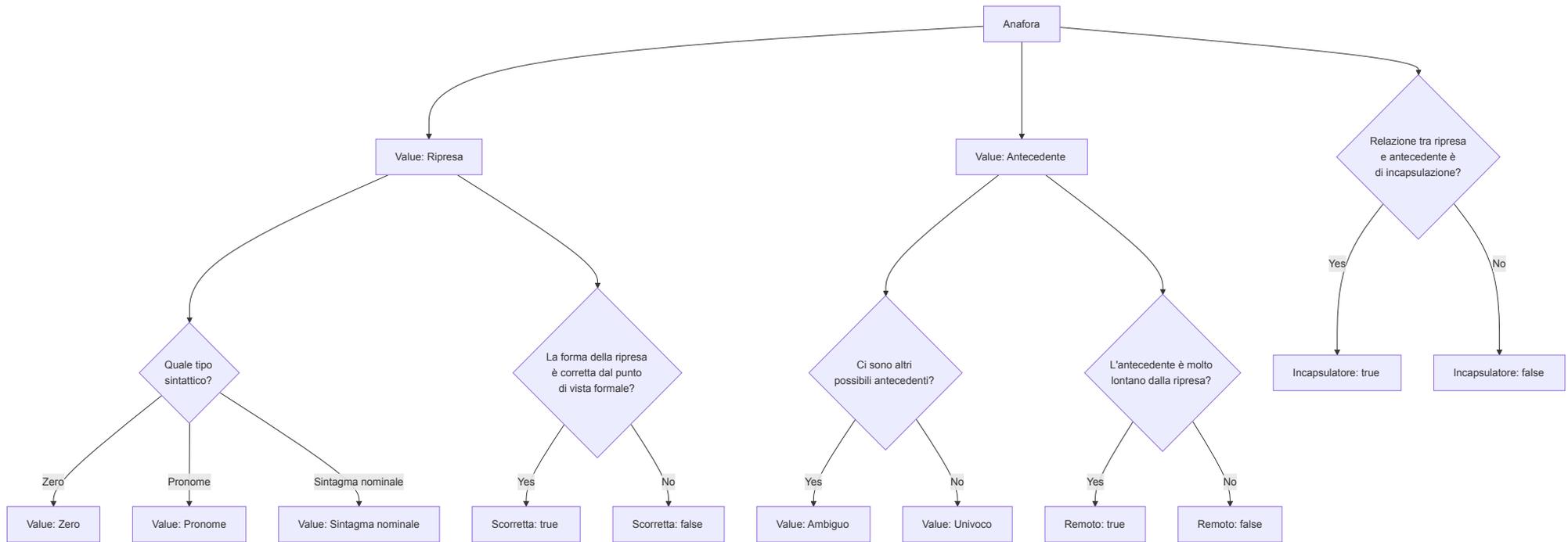
**Node annotation:** AnAntecedenteRemoto; **Value:** true, false

**Edge type:** Anafora

**Edge annotation:** Incapsulatore; **Value:** true, false

Definizione "anafora problematica": difficoltà o impossibilità di interpretare alcuni sintagmi nel testo che si riferiscono a un altro costituente che compare nella parte precedente o successiva del testo stesso. (Esempi di anafore problematiche compaiono in: De Masi, Maggio (2008); Prada (2009); Serianni, Benedetti (2009); Roggia (2010); Colombo (2011); Ruggiano (2011); Palermo (2013); Notarbartolo (2014); Tarallo (2019).

Procedura di annotazione:



1. segnalare se l'elemento annotato è una ripresa o un antecedente/sussequente.
2. se è una ripresa, segnalarne il tipo sintattico, se pronominale, zero o sintagma nominale.
3. se è una ripresa, segnalare se la forma della ripresa è corretta dal punto di vista formale.
4. se è un antecedente/sussequente, segnalare se è ambiguo (ci sono altri possibili referenti?) o se è univoco;
5. se è un antecedente/sussequente, segnalare se è remoto (è molto lontano dalla ripresa?);
6. nel caso in cui sia possibile identificare un antecedente/sussequente, tirare l'arco dalla ripresa all'antecedente/sussequente (SEMPRE IN QUESTA DIREZIONE) e segnalare se la relazione anaforica è tra due entità singole (es. tra SN e pronome, zero, SN) oppure tra una porzione più ampia di testo e la ripresa (es. tra frase e pronome, zero o SN). In questo caso la ripresa è un incapsulatore.

Definizione **anafora pronominale**: possono fungere da collegamenti anaforici i pronomi:

- Personali, dimostrativi, relativi.
- Avverbi pronominali: qui, lì, così con valore locativo e modale.
- Modificatori possessivi
- Pronomi indefiniti: entrambi, l'uno e l'altro, tutti (quando si riferisce alla totalità degli elementi denotati dall'antecedente).

Esempi di pronomi problematici:

Come confermato da un rapporto ISTAT del 2021, **ne** sono derivati molti danni alle vite degli studenti (...). (referente ambiguo o non ricostruibile)

La referenza pronominale, in particolare, può avere (anche) problemi formali: usare l'attributo AnRipresaScorretta = "true" quando la forma pronominale è errata dal punto di vista formale.

*ciò che il professore può vedere è solo ciò che lo studente vuole farli (fargli) vedere* (Corpus ITACA)

Particolare attenzione a:

- pronomi soggetto (egli, ella, esso, essa, essi, esse) usati nei casi obliqui: es. *a egli è venuto in mente di partire per la Francia* (al posto di *a lui è venuto in mente di...*) (Esempio fittizio);
- alla forma di cortesia (incluse le riprese pronominali: confusione tra il 'lei' di cortesia e il 'voi');
- al pronome indiretto femminile "le" = a lei (spesso assorbito da "gli" v. neostandard);

#### Warning

ATTENZIONE: non annotare i tratti neostandard "lui/lei" al posto di "egli/ella"; "gli" al posto di "loro"

- ai pronomi relativi

*Ecco perché ci tengo che Le arrivi questa lettera, dove cercherò di spiegare e analizzare i motivi **per il quale** (per i quali) la DDI sta causando un declino del benessere mentale dei miei coetanei fin troppo velocemente.* (Corpus ITACA)

#### Warning

ATTENZIONE: le devianze formali sui pronomi relativi si intrecciano alle tendenze di ristandardizzazione del paradigma dei pronomi relativi. In particolare non segnaliamo il 'dove' non locativo, a meno che non ci siano problemi di ambiguità del riferimento, come accade in:

"La scuola costituisce il momento di maggior socialità per una grande parte dei ragazzi, dove si instaurano rapporti e dove si è continuamente inseriti in un gruppo".  
(Corpus ITACA)

Commento: 'dove' ambiguo tra "la scuola" e "il momento").

Definizione **anafora zero**: possibilità delle lingue pro-drop di non esprimere linguisticamente il soggetto. Le regole per la possibilità di non espressione linguistica del soggetto sono estremamente complesse e sono difficilmente formalizzabili. Esse comprendono la natura argomentale del verbo (se transitivo o ditransitivo), la natura della

relazione discorsiva tra frasi adiacenti, la salienza dell'antecedente che deve essere ripreso, ecc.

Casi speciali: abbiamo annotato sotto il caso di "anafora zero" anche i casi di **gerundi sospesi**, ovvero:

- gerundi che modificano un elemento nominale e quindi hanno funzione attributiva, tanto che si possono sostituire con una frase relativa;
- gerundi che hanno funzione predicativa, tanto che possono essere sostituiti con una frase pseudo-relativa (Cinque 1989, p. 501) o con l'infinito o con una temporale;
- casi in cui nella principale si usi la diatesi passiva e uno stile nominale. Il gerundio, soprattutto se posposto, può generare ambiguità di coreferenza tra il soggetto sintattico e quello logico (e quindi vengono segnalati come antecedenti ambigui);  
si veda l'opinione dell'[Accademia della Crusca](#) sull'argomento.

Esempi di anafore zero problematiche:

*riuscì (scil.: Ulisse) ad un tratto a spingere il palo d'ulivo e (XX lui e i compagni?) lo **ficcarono** dentro l'occhio del Ciclope.* (Ruggiano 2011: 132).

*L'aspetto, che secondo me è più di altri fondamentale per il mio antagonismo verso la DDI, è che (XX la DDI/questa) **causa** la fusione del luogo dello studio con quello che è invece il proprio spazio privato.* (Corpus ITACA)

Definizione **incapsulatore**: sintagmi nominali o pronomi (es. *ciò*) che riprendono il contenuto di un enunciato o una sequenza di enunciati.

Esempio di incapsulatore problematico:

*Per prima cosa non si instaurerebbero più legami tra i professori e gli studenti, o comunque il rapporto non crescerebbe vista la mancanza fisica, **il che** ritengo sia fondamentale durante un percorso scolastico per vivere bene l'ambiente in cui ci si trova.* (Corpus ITACA)

Commento: l'incapsulatore incapsula elementi semanticamente antitetici rispetto a quello che lo scrivente aveva in mente (l'instaurarsi di legami tra professori e studenti, il crescere del rapporto vista la vicinanza fisica).

## Dispositio

**Node annotation:** DisposizioneErrata; **Value:** Spostare, Togliere

Definizione di disposizione errata degli argomenti: introduzione e sviluppo di un'informazione nuova; poi, mediante una progressione tematica, si introduce un'altra informazione nuova e da ultimo si riprende la prima informazione (ormai diventata data) con conseguente interruzione della progressione tematica. (Serianni, Benedetti 2009)

- usare l'attributo "Spostare" nei casi in cui la sequenza testuale non si colleghi a livello logico, tematico o temporale con le sequenze adiacenti.

*grazie alla dad abbiamo più tempo per noi e possiamo vivere più tranquillamente senza fare tutto di corsa. **Abbiamo migliorato le nostre competenze tecnologiche** ed inoltre permette di poter fare le cose personali come fare colazione con più calma potendo iniziare il lavoro o la scuola semplicemente accedendo il computer e andando online.*

Tematicamente, la sequenza è unitaria (con la dad possiamo fare le cose più con calma) se non fosse per l'inserzione sulle competenze tecnologiche che vengono comunque riprese nel seguito del testo.

- usare l'attributo "Togliere" nei casi in cui la sequenza testuale non si collega a livello logico, tematico o temporale con nessuna delle sequenze del testo.

NB. spesso ci sono sequenze testuali interpolate ad altre che bloccano la progressione tematica e non sono collegate a nessun'altra parte di testo.

*Diverse scuole non hanno applicato dei provvedimenti che comprendessero un'innovazione didattica in modo da permetterci di tornare in classe in sicurezza. **Circa il 62 % ritiene che la lezione frontale sia stata la prassi più comune nella propria scuola , questo è stato scritto dall' A. S. nel 2021. Nella mia scuola l' unico provvedimento che è stato messo in atto è stato cambiare la disposizione delle classi in base al numero dei componenti in tale e alla grandezza dell' aula.***

Ad esempio, in questo caso lo studente parla dei provvedimenti presi per tornare a scuola in sicurezza. Interpola una sequenza sul tipo di didattica applicata in DaD (forse collegata a "innovazione didattica") e poi torna a parlare dei provvedimenti per tornare a scuola in sicurezza.

NB. spesso ci sono sequenze testuali, di solito poste dopo le conclusioni, e che sembrano rimanere "appese" o non ben tematicamente integrate rispetto al resto del testo (che già sembrava essersi concluso con le conclusioni). Consideriamo questi casi come off-topic (segnaliamo con l'attributo "Togliere").

*Spero che le mie considerazioni possano farla riflettere sul fatto che ognuno di noi è diverso e nell'apprendere non vi è nulla di sbagliato si possono usare vari metodi e sperimentare nuove tecniche per l'apprendimento, basta che funzionino.*

***Spesso mi domando quanto a lungo mi rimarrà impresso ciò che ho studiato il giorno precedente o gli anni prima, la mia risposta è sempre la stessa.***

***Ciò che rimane impresso è ciò che abbiamo capito non imparato a memoria, è ciò che ci è stato spiegato con nuove metodologie di apprendimento, qualcosa che catturi l'attenzione e ci faccia venire la voglia di approfondire nuovi aspetti per conto nostro.***

## Informazione nuova non sviluppata

**Node annotation:** RemaNonEspanso

Definizione informazione nuova non sviluppata: l'informazione nuova introdotta nel testo non viene ripresa in alcun modo (né diretto, né associativo) nel o negli enunciati successivi. (in Notarbartolo, 2014 il fenomeno è riportato sotto il nome di "rema non espanso" o "promessa non mantenuta").

Usare il tag "RemaNonEspanso" marcando la porzione di testo che corrisponde all'enunciato che contiene l'informazione nuova, di norma coincidente con la frase (da punto a punto o da punto a virgola splice).

*Avendo vissuto circa due anni di lezioni a distanza, **posso sostenere di aver compreso e ben individuato quali possono essere i vantaggi e gli svantaggi di questo sistema scolastico. Personalmente ritengo che la Didattica a distanza possa essere veramente utile in termini di organizzazione dello studente, ma la sua efficacia dipende da fattori difficili da gestire.*** (Corpus ITACA)

Commento: *i vantaggi e gli svantaggi* non vengono più ripresi/menzionati in maniera lineare.

Le sue prolusioni offrono spesso molti spunti di riflessione (ndr: si sta parlando di un ecclesiastico). È stato il caso anche dell'ultima (...). **In questa sede ci si limiterà a segnalare qualche passaggio particolarmente stimolante.** Ce ne sarebbero anche altri: per esempio, la forse valorizzazione della pietà popolare, oppure la considerazione sulla crisi della fede in Occidente, sullo sforzo della chiesa in ambito caritativo, sulle "nuove malattie" della modernità. (Notarbartolo 2014: 125).

## Reading protocol

To guide annotators in the task of annotating the textual features listed above, it has been proposed to them to follow a reading protocol. This comprises 5 steps, from pseudonymization to the annotation of textual phenomena on the macro-level. During Step 0 (Pseudonymization) and Step 3 (Textual structure) new annotations were produced, which are also searchable in the annotated corpus. For these two steps we provide detailed guidelines. For step 1 (Segmentazione), 2 (Textual features micro), and 4 (Textual features macro), we refer to the annotation schema detailed above.

### 0. Pseudonymization

**Node annotation:** Pseudonymization; **Value:** Forename, Surname, Animalname, Placename, Schoolname, Classname, Other

#### Note

These guidelines are adapted from the pseudonymization guidelines used for the LEONIDE project, available at [this link](#)

Use this annotation layer for real person's names (e.g. teachers of the class, students), names of existing animals, place names within the area of South Tyrol, and names of schools.

Regarding person names: Write "Forename" and "Surname" respectively instead of the real names used in the original text, and use the tags for each token (i.e. use one tag for the first name and another one for the surname). For initials just use the tag "Surname" for the whole chunk (e.g. *WNC*)

Regarding names of animals: Write "Animalname" instead of the real name of an animal (dog, horse), and use one tag even if the name consists of two tokens.

Regarding names of places: Write "Placename" instead of the real names of cities, places and streets, or other names that refer to real existing places in the environment of the writers (e.g. bars). Use one tag for each place.

Regarding names of schools: Write "Schoolname" instead of the real name of the school. Anonymise also commonly used abbreviations that are linked to a specific school. Use one tag even if the real name of the school consists of two + tokens.

Regarding the names of the classes: Write "Classname" instead of the real name of the class. Use one tag per class.

Do not anonymize invented names or place names in a generic usage (e.g. *It is important to know Italian in Bolzano, Merano, Bressanone but in the valleys of South Tyrol it is more important to know German*) but only cases when it potentially helps to identify real persons (e.g. *I live in PLACENAME.*) or schools (e.g. *after the middle school*

in PLACENAME, I will go to the SCHOOLNAME in PLACENAME).

## 1. Segmentazione

Annotate all linguistic phenomena listed under "Segmentazione".

## 2. Fenomeni testuali micro

Annotate all linguistic phenomena related to connettivi e anafore.

## 3. Struttura testuale

**Node annotations:** text\_structure; **Values:** Intro, Tesi, Argopro, Argocontro, Temp, Concessione, Confutazione, Digr, Concl, Other

**Related annotations:** text\_structure\_comment

### Note

These guidelines are translated and adapted from those used for the KoKo project (Abel et al. 2014), unpublished.

Annotazioni propedeutiche alla segnalazione di fenomeni testuali macro. Lettura che serve a suddividere il testo nelle sue componenti principali, quali:

- Introduzione (Intro): la funzione comunicativa di un'introduzione è di preparare il destinatario a ricevere il testo. Dichiarando l'argomento, l'occasione della scrittura, la funzione del testo (ad esempio, persuadere/convincere), il destinatario ha l'opportunità di attivare conoscenze precedenti o aspettative specifiche sulla sequenza dei contenuti come base per la comprensione delle informazioni contenute nel corpo principale. In questo senso un'introduzione può facilitare e guidare la ricezione (Jechle 1992: 139).
- Tesi (Tesi): la tesi è la posizione che l'autore sceglie di sostenere con ragioni/prove a supporto. Più in generale, la tesi è la proposizione «della cui verità si vuole convincere l'ascoltatore» (Santambrogio 2006: 35).
- Argomenti a favore (Argopro): tutto ciò che si scrive con l'intento di rendere una tesi più credibile di quanto essa non fosse in assenza dell'argomentazione (Santambrogio, 2006: 61). Se nel testo non è presente una tesi si usa il tag "Argopro" per segnalare i vantaggi.
- Argomenti a sfavore (Argocontro): tutto ciò che si scrive nell'ottica di includere affermazioni che indeboliscono la credibilità della tesi. Se nel testo non è presente una tesi, si usa il tag "Argocontro" per segnalare gli svantaggi.

### Warning

ATTENZIONE: Il carattere pro o contro degli argomenti è sempre legato alla tesi dello scrittore.

- Argomentazione concessiva: solitamente, l'argomentazione concessiva si trova in corrispondenza dell'introduzione di argomenti a sfavore della tesi, in quanto è la presa e il disinnescamento di un argomento opposto. Argomentando concessivamente, cioè riprendendo i controargomenti e mostrando che non indeboliscono la sua posizione, lo scrivente può dare al suo testo un maggior grado di plausibilità (Steinhoff 2007: 329). Componenti dell'argomentazione concessiva sono:
  - Concessione: concessione alla correttezza dell'argomentazione dell'avversario;
  - Confutazione: invalida l'argomentazione dell'avversario (argomenti a sfavore) e sostiene la tesi dell'autore.
- Sequenza temporale (Temp): tutto ciò che si scrive con l'intento di collocare nello spazio e nel tempo avvenimenti inerenti al tema trattato.
- Digressione (Digr): sequenze tematiche che si allontanano troppo dal nucleo tematico trattato nel resto del testo.
- Conclusioni (Concl): una conclusione è un breve e conciso riassunto delle affermazioni centrali del testo. La conclusione dovrebbe riferirsi a ciò che è stato scritto o dovrebbe essere una conclusione che può essere derivata da ciò che è stato scritto.

 Tip

## 4. Fenomeni testuali macro

Annotate all linguistic phenomena related to contraddizione, informazione nuova non sviluppata, dispositio.

---

## References

Abel A., Blaschitz V., Glaznieks A. (2014). Erläuterungen zu den Analyse Kriterien – Bereich Textqualität, Projekt KoKo, unpublished.

Ciccolone, S. (2012). Incoerenze testuali e problemi di combinazione lessicale nella produzione scritta di studenti universitari: primi rilievi e proposte esplicative. In G. Bernini, C. Lavinio, A. Valentini, & M. Voghera (Eds.), *Atti dell'XI Congresso dell'Associazione Italiana di Linguistica Applicata. Bergamo, 9-11 giugno 2011*. Guerra Edizioni.

Cignetti, L. (2015). Didattica della scrittura e linguistica del testo: tre priorità di intervento. In M. Ostinelli (Ed.), *La didattica dell'italiano. Problemi e prospettive* (pp. 14–24). SUPSI.

Colombo, A. (2011). *A me mi: dubbi, errori, correzioni nell'italiano scritto*. FrancoAngeli.

De Masi, S., & Maggio, M. (2008). *Pratiche di scrittura nella scuola superiore*. FrancoAngeli.

- De Santis, C., & Gatta, F. (2013). Notizie dalla scuola. Le competenze grammaticali e testuali degli studenti madrelingua all'uscita della scuola secondaria. Risultati di un'indagine. *Studi Di Grammatica Italiana, XXXI–XXXII*, 411–441.
- Ferrari, A. (2010). connettivi. In *Enciclopedia dell'Italiano*. Treccani. [https://www.treccani.it/enciclopedia/connettivi\\_\(Enciclopedia-dell'Italiano\)](https://www.treccani.it/enciclopedia/connettivi_(Enciclopedia-dell'Italiano))
- Ferrari, A. (2014). *Linguistica del testo. Principi, fenomeni, strutture*. Carocci.
- Ferrari, A. (2021). Segnali discorsivi e connettivi. *Lingua e Stile, LVI*(1), 143–150. <https://doi.org/10.1417/100904>
- Ferrari, A., Lala, L., Longo, F., Pecorari, F., Rosi, B., & Stojmenova Weber, R. (2018). *La punteggiatura italiana contemporanea. Un'analisi comunicativo testuale*. Carocci.
- Ferrari, A., Lala, L., & Zampese, L. (2021). *Le strutture del testo scritto. Teoria e esercizi*. Carocci.
- INVALSI. (2012). *Rilevazione degli errori più diffusi nella padronanza della lingua italiana nella prima prova di italiano*.
- Jechle, T. (1992): *Kommunikatives Schreiben: Prozeß und Entwicklung aus der Sicht kognitiver Schreibforschung*. Narr.
- Notarbartolo, D. (2014). *Competenze testuali per la scuola*. Carocci.
- Palermo, M. (2013). *Linguistica testuale dell'italiano*. il Mulino.
- Perelman, C., & Olbrechts-Tyteca, L. (1966). *Trattato dell'argomentazione. La nuova retorica*. Einaudi.
- Prada, M. (2009). Le competenze di scrittura e le interazioni comunicative attraverso lo scritto: problemi e prospettive per una didattica della scrittura. *Italiano LinguaDue, 1*(1), 232–278.
- Prada, M. (2016). Scritto e parlato, il parlato nello scritto. Per una didattica della consapevolezza diamesica. *Italiano LinguaDue, 2*, 232–260.
- Roggia, C. E. (2010). Una varietà dell'italiano tra scritto e parlato: la scrittura degli apprendenti. In A. Ferrari & A.-M. De Cesare (Eds.), *Il parlato nella scrittura odierna. Riflessioni in prospettiva testuale* (pp. 197–224). Peter Lang.
- Ruele, M., & Zuin, E. (2020). *Come cambia la scrittura a scuola*. IPRASE.
- Ruggiano, F. (2011). *L'italiano scritto a scuola*. Aracne.
- Santambrogio, M. (2006). *Manuale di scrittura (non creativa)*. Laterza.
- Serianni, L. (2009). *Scritti sui banchi: l'italiano a scuola tra alunni e insegnanti*. Carocci.
- Steinhoff, T. (2007): *Wissenschaftliche Textkompetenz*. Niemeyer.

Tarallo, C. (2019). Distribuzione pentenaria e analisi della competenza testuale su gruppi differenziali di studenti. In M. Palermo & E. Salvatore (Eds.), *Scrivere nella scuola d'oggi*. Franco Cesati Editore.

## Progetto Itaca

### Item della scala e relativa scala di risposta

Dimensione	ITEM	NOTA	TESTO	Scala di risposta (grado di accordo)
<b>D1 STRUTTURAZIONE DEL TESTO (ARGOMENTATIVO)</b>	I024		La tesi è sostenuta da argomenti	0=Del tutto in disaccordo; 1=Un po' in disaccordo; 2=Un po' d'accordo; 3=Del tutto d'accordo
	I026		Le conclusioni derivano dagli argomenti presenti nel testo	0=Del tutto in disaccordo; 1=Un po' in disaccordo; 2=Un po' d'accordo; 3=Del tutto d'accordo
	I031		Il testo è stato progettato in modo efficace	0=Del tutto in disaccordo; 1=Un po' in disaccordo; 2=Un po' d'accordo; 3=Del tutto d'accordo
	I022		Il testo è strutturato sulla base dell'idea di fondo	0=Del tutto in disaccordo; 1=Un po' in disaccordo; 2=Un po' d'accordo; 3=Del tutto d'accordo
	I034		Gli argomenti introdotti nel testo sono funzionali allo scopo	0=Del tutto in disaccordo; 1=Un po' in disaccordo; 2=Un po' d'accordo; 3=Del tutto d'accordo
<b>D2 COMPRESIBILITÀ</b>	I037	REVERSE	I termini utilizzati provocano fraintendimenti	3=Del tutto in disaccordo; 2=Un po' in disaccordo; 1=Un po' d'accordo; 0=Del tutto d'accordo
	I017		Il collegamento tra la ripresa e il suo referente è sempre ricostruibile	0=Del tutto in disaccordo; 1=Un po' in disaccordo; 2=Un po' d'accordo; 3=Del tutto d'accordo
	I043		Il testo si comprende facilmente	0=Del tutto in disaccordo; 1=Un po' in disaccordo; 2=Un po' d'accordo; 3=Del tutto d'accordo
	I040		Tutte le informazioni necessarie alla piena comprensione del testo sono ben esplicitate	0=Del tutto in disaccordo; 1=Un po' in disaccordo; 2=Un po' d'accordo; 3=Del tutto d'accordo
	I020		L'uso dei connettivi e delle anafore è funzionale alla coesione del testo	0=Del tutto in disaccordo; 1=Un po' in disaccordo; 2=Un po' d'accordo; 3=Del tutto d'accordo
<b>D3 SEGMENTAZIONE</b>	I013		Nel testo vengono rispettate le regole di interpunzione	0=Del tutto in disaccordo; 1=Un po' in disaccordo; 2=Un po' d'accordo; 3=Del tutto d'accordo
	I012	REVERSE	Le devianze nella punteggiatura sono tali da compromettere la coerenza e/o la comprensione del testo	3=Del tutto in disaccordo; 2=Un po' in disaccordo; 1=Un po' d'accordo; 0=Del tutto d'accordo
	I005	REVERSE	All'interno del paragrafo la punteggiatura è scarsamente segnalata	3=Del tutto in disaccordo; 2=Un po' in disaccordo; 1=Un po' d'accordo; 0=Del tutto d'accordo
	I006		L'uso del punto nel testo è sempre funzionale alla suddivisione degli enunciati	0=Del tutto in disaccordo; 1=Un po' in disaccordo; 2=Un po' d'accordo; 3=Del tutto d'accordo
	I007		Gli incisi sono di norma correttamente segnalati	0=Del tutto in disaccordo; 1=Un po' in disaccordo; 2=Un po' d'accordo; 3=Del tutto d'accordo

## Tabella di conversione dei punteggi grezzi in punteggi su scala ad intervalli (logit)

Dimensione	Punteggio grezzo	Punteggio logit
<b>D1 STRUTTURAZIONE DEL TESTO (ARGOMENTATIVO)</b>	0	1
	1	1.926
	2	2.782
	3	3.351
	4	3.800
	5	4.189
	6	4.547
	7	4.896
	8	5.252
	9	5.634
	10	6.061
	11	6.555
	12	7.143
	13	7.875
	14	8.902
	15	10
<b>D2 COMPRESIBILITÀ</b>	0	1
	1	2.042
	2	3.003
	3	3.632
	4	4.125
	5	4.544
	6	4.921
	7	5.276
	8	5.624
	9	5.980
	10	6.359
	11	6.783
	12	7.285
	13	7.933
	14	8.925
	15	10
<b>D3 SEGMENTAZIONE</b>	0	1
	1	1.918
	2	2.773
	3	3.362
	4	3.839
	5	4.250
	6	4.618
	7	4.959
	8	5.288
	9	5.621
	10	5.975
	11	6.377
	12	6.870
	13	7.558
	14	8.722
	15	10

## Come si usa la scala?

### **PRIMO STEP: la valutazione per singoli item**

Valutare il testo su ciascun **item** usando la **scala di risposta** (p. 1), che indica il grado di accordo o disaccordo del valutatore rispetto a ciascuna affermazione nella scala. Es. se il valutatore è del tutto d'accordo con l'affermazione "La tesi è sostenuta da argomenti", allora assegnerà il punteggio di 3 a questo item. Ripetere la procedura per tutti gli item.

**ATTENZIONE!** Gli item "reverse" vanno intesi al contrario: se si è del tutto d'accordo con un'affermazione negativa come "All'interno del paragrafo la punteggiatura è scarsamente segnalata", allora si assegnerà il punteggio 0; un po' d'accordo = 1; un po' in disaccordo = 2; del tutto in disaccordo = 3. Vedi item I037, I012, I005.

### **SECONDO STEP: la somma per dimensioni**

Alla fine della valutazione, il testo avrà ricevuto un punteggio da 0 a 3 per tutti gli item. Sommare i punteggi per ciascuna dimensione. Es. se per la dimensione di "strutturazione del testo (argomentativo)", il testo ha ottenuto i punteggi 3, 2, 1, 2, 2, avrà ottenuto un punteggio grezzo di 10. Ma non è finita qui!

### **TERZO STEP: conversione alla scala logit**

Usare la **tabella di conversione dei punteggi grezzi in punteggi su scala ad intervalli (logit)** (p. 2). Es. se per la dimensione 1 "strutturazione del testo argomentativo" un testo ottiene un punteggio grezzo di 10 (ossia la somma degli item della dimensione 1 dà valore 10) il suo punteggio logit sarà 6.061. Se per la dimensione 3 "segmentazione" si ottiene, per esempio, un punteggio grezzo di 5 (ossia la somma degli item della dimensione 3 dà valore 5), il suo punteggio logit sarà 4.250.

**ATTENZIONE!** Nota che le corrispondenze tra punteggio grezzo e scala logit sono leggermente diverse per ciascuna dimensione; questo avviene perché la scala logit "standardizza" i risultati sulla base della difficoltà dei singoli item. Può essere che in "strutturazione del testo argomentativo" ci siano item più difficili che in "segmentazione" (ovvero, globalmente, nel nostro dataset di validazione della scala, meno testi hanno ricevuto punteggi alti negli item di "strutturazione"). La scala logit tiene conto di queste differenze e rende comparabili le tre dimensioni.

### **QUARTO STEP: interpretare il numero**

La scala logit restituisce un valore complessivo per ciascuna delle dimensioni testuali che abbiamo individuato: "strutturazione del testo", "comprensibilità" e "segmentazione". Ma come va interpretato? Più è alto il valore logit e più il testo possiede quella caratteristica (è ben strutturato, è comprensibile, è ben segmentato). Il mio consiglio è quello di usare questo strumento come strumento diagnostico e non valutativo! Lo strumento restituisce un'immagine puntuale di certe caratteristiche testuali, ma non può sostituire la valutazione fatta dal/la docente, la quale tiene in considerazione anche altri fattori, quali i progressi del singolo studente o studentessa, globali o locali. Inoltre, la scala da 1 a 10 non può essere intesa come mero sostituto del voto! Il 4, ad esempio, non equivale al 4 scolastico. Esso indica semplicemente che la caratteristica indagata in quel testo è presente in maniera leggermente al di sotto della media.

Happy rating!