

Alma Mater Studiorum Università di Bologna

DIPARTIMENTO DI INTERPRETAZIONE E TRADUZIONE

Corso di Laurea magistrale Specialized Translation (classe LM - 94)

TESI DI LAUREA

in Language Data Analysis

Italian word lists and academic language: a corpus-based study of student writing

CANDIDATA:

Alessandra Pierantoni

RELATRICE:

Maja Miličević Petrović

CORRELATRICE:

Silvia Bernardini

Anno Accademico 2022/2023
Secondo Appello

*A chi conosce il potere delle parole,
e a chi non lo conosce ancora.*

*To those who know the power of words,
and to those who do not know it yet.*

Acknowledgements

Even though I grew a lot in these two years of Master's and I have been working on myself for quite some time now, I still would not have reached this milestone if it was not for the people that accompanied me in this long, difficult but also exciting journey.

First of all, I would like to express my gratitude to Professor Maja Miličević Petrović for not only supporting my work, but also for offering her help and for listening to me when I was struggling. It is not common, unfortunately, for us students to be asked how we are feeling.

I would also like to thank Professor Silvia Bernardini for her useful inputs on my work and for her quick responses.

Thanks to Lorenzo Zanasi, Arianna Bienati, and the entire team behind ITACA and the Institute for Applied Linguistics at Eurac Research for welcoming me one year and a half ago, for being so passionate about your work and for making me discover the world of research. I still remember the emotion I felt when reading ITACA's description on Eurac website; it felt like it was calling for me, and it makes sense since I love writing and the first language I fell in love with was Italian. This dissertation would not exist if I did not get to know you and your amazing projects. A special thanks also to Stefania Spina for helping me with her word list.

Thanks to Rachele and Francesca, who I had the incredible luck to meet two years ago, for helping me stand up every time I fall and for bringing the light and lucidity in those moments of chaos that I thought would have lasted forever; this friendship is without a doubt the most beautiful thing that these two years in Forlì gave me.

Thanks to Virginia and Paolo for listening to my embarrassingly long stories and for making me laugh every time.

Thanks to Ilaria for being one of the first people with whom I felt at ease in the first year of master's. Thanks to Alessia for being the only student, besides me, to have the magic linguistic combination French-German and therefore for being an incredible source of support and wisdom, and also a great company during lunch breaks (with mostly vegan food).

Thanks to Giulia, Rossella and Clara for staying with me even after our paths divided. My true life in Forlì started after meeting you, perhaps, and even if our lives took us to different places, I know I can always reach out to you to ask you for advice, share amazing songs and talk about how small the world actually is.

Thanks to Agnese and Giacomo for making me live experiences I had never thought I would have done and for believing in me when I could not really see what I was capable of. I feel so honored and lucky for having had you by my side in the last year. I would simply not be who I am today if I did

not meet you. My journey in the world of activism, which I hope will continue, will not be the same without you.

Thanks to Auri for the sensitivity and the incredible tact. It was not taken for granted, and I am happy to have witnessed these qualities of yours before leaving Forlì.

Thanks to Nadine and Donatella for supporting me when I wanted to make my voice heard, when I just needed some comfort and obviously for fighting for our future: you are truly an inspiration.

Thanks to Carmela, Rebecca and all the cool people behind Spazio 2030 for giving me the opportunity to discover a very important part of the world, which eventually became a big part of myself, too. It was the start of something beautiful that changed my life completely.

Thanks to Caterina and Gaia for reminding me that there is always a solution, even when dealing with the (occasional) brutality of university.

Thanks to Schegge for being a safe space where I could start writing again. I found again a piece of me with this opportunity. I am sure you will continue to do amazing work. A special thanks to Caterina, who was one of the first people I talked to about activism and introduced me to some of the considerations I still bring with me today.

Thanks to Hikma, especially the communication team, for being so awesome and supportive. I will never forget this year's summit: it was brief, but intense, as they say.

Thanks to all the other incredible people I met in the world of activism that made me feel at home, each in their own way, no matter where I was or how well we knew each other, and who I still remember very fondly today: I will continue to cherish every conversation, every smile and every hug. I hope to see you again soon.

Thanks to Forlì too, of course. It took me some time to love you, but eventually you became one of my homes. You made me grow a lot and you helped me express myself a lot better. I finally got to let my voice out here, after spending years seeing a louder and braver version of myself only inside of me.

Last but not least, thanks to my family for supporting me and my choices from the start.

Abstract

English: Analyzing vocabulary choices made by students for their writing tasks is important to better understand the way young pupils express themselves in academic contexts. Studies of academic language based on corpora or on word lists have been conducted for many years now, but despite this, there are not many first language (L1) student corpora focusing on academic writing, and word lists are primarily used for analyzing second language (L2) vocabulary. This is the one of the first empirical studies on an L1 Italian corpus of student writing which aims to explore how relevant Italian word lists are for texts written by native speakers, in particular if the words present in the lists are representative of the way young people write. The analyses conducted for this research pave the way for more in-depth examination of word lists, especially for L1 native speakers, and of Italian academic language knowledge and acquisition in high school. Using software for vocabulary profiling and corpus analysis, the potential of word lists will be discussed, to see if they can also be used as a learning tool by L1 students and not just by L2 learners, who were the original target of this type of resource. An exploration of the possibilities offered by using of word lists could also have implications for teaching and for the research focusing on young people and academic language.

Italiano: Analizzare le scelte lessicali compiute dagli studenti per i loro testi scritti è importante per poter comprendere meglio il modo in cui i giovani si esprimono nei contesti accademici. Gli studi sul linguaggio accademico basati sui corpora o sulle liste di parole hanno una lunga storia; ciononostante, non ci sono molti corpora che raccolgono testi scritti da studenti nella loro prima lingua (L1) che abbiano come focus la scrittura accademica, e le liste di parole vengono usate principalmente per analizzare il lessico della seconda lingua (L2). Questo è uno dei primi studi sperimentali su un corpus in italiano L1 di testi scritti da studenti che punta a constatare la rilevanza delle liste di parole italiane per i testi scritti da parlanti nativi; in particolare l'obiettivo è stabilire se le parole presenti nelle liste sono rappresentative del modo in cui i giovani scrivono. Le analisi condotte per questa ricerca fanno da apripista per analisi ancora più approfondite delle liste di parole, soprattutto nel contesto dei parlanti L1, e della conoscenza e dell'acquisizione del linguaggio accademico italiano nelle scuole superiori. Tramite l'uso di software per la profilazione del lessico e per l'analisi di corpus sarà discusso il potenziale delle liste di parole, per verificare se queste possono essere usate come strumento di apprendimento dagli studenti L1 e non solo dagli apprendenti L2, che costituiscono il target originale di questa risorsa. Un'esplorazione delle possibilità offerte dall'uso delle liste di parole potrebbe anche avere implicazioni sull'insegnamento e sulla ricerca incentrata sul rapporto tra i giovani e il linguaggio accademico.

Table of contents

Introduction	10
1 Academic language.....	12
1.1 Definitions of academic language	12
1.2 Main characteristics of academic language.....	14
1.3 Acquisition and teaching of academic language for L1 and L2	17
1.4 Academic Italian vs. academic English.....	19
1.5 Previous corpus-based studies on learners' academic language.....	20
1.5.1 Previous studies on L1 learners' academic language.....	21
1.5.2 Previous studies on L2 learners' academic language.....	22
2 Vocabulary and word lists.....	24
2.1 Definitions and types of vocabulary.....	24
2.2 Vocabulary knowledge and acquisition	26
2.2.1 Vocabulary and L1	28
2.2.2 Vocabulary and L2	29
2.3 Word Lists	30
2.3.1 English word lists	30
2.3.2 Italian word lists	31
2.4 Studies based on word lists	33
2.4.1 Use of word lists for teaching.....	34
2.4.2 Problems with word lists	34
3 Empirical study.....	37
3.1 Research question.....	37
3.2 Method.....	37
3.2.1 The corpus.....	37
3.2.2 Data and analyses	38
3.3 Results	40
3.4 Discussion	46
Conclusion.....	48
References	50
Appendix	59

Introduction

As we know, language is strongly connected with society (Kadarisman, 2009); therefore, I believe that studying the way young people write today is fundamental in order to better understand our fast-changing world. There are different types of writing that we see in our everyday lives: narrative, in the books and short stories we read; expository, in textbooks and instructions; descriptive, in memoirs and poetry; persuasive, in marketing and editorial newspaper articles; with the rise of social media and messaging apps, a lot of attention started being given to informal shorter forms of writing, too. Given the amount of time spent in formal education (school and possibly university), one of the most important types of written language is the academic one.

Academic language is a dynamic study field which focuses on the language used in education contexts. Among its subtopics we find word lists, i.e., lists of words considered as useful for academic writing; such lists have been a research topic in this area for many decades now, along with studies based on collections of academic language use – text corpora. Research, however, has been focusing mainly on vocabulary and word lists in a second language (L2); to contribute to filling the gap in research on first language (L1) vocabulary development, I decided to conduct an analysis on the vocabulary choices made by L1 Italian students, to see if existing Italian word lists can be applied also to profile their vocabulary. To do this, I first reviewed existing vocabulary studies focused on corpora and word lists; then, I analyzed an L1 Italian corpus of student writing with a focus on the relevance of Italian word lists.

I decided to work on this topic after having done my curricular internship at the Institute for Applied Linguistics at Eurac Research (Bolzano). There, I discovered the world of research, and I worked on the ITACA project, one of whose results is the corpus I used for this study. Reading the texts written by students that were collected for the project made me curious about the students' vocabulary choices.

In this thesis, I will first give an overview of academic language, its theoretical background, its characteristics, a comparison between academic Italian and English and the studies that have already been published (concerning both of these languages, but with a focus on Italian). In Chapter 2, I will present work on vocabulary, its knowledge and acquisition both for Italian and English, and describe some word lists, including the ones I used for the analysis, with a presentation of their role for teaching purposes and a description of some of the common issues associated with them. Finally, Chapter 3 will be focused on the empirical study and the analyses I conducted, and it will also present a discussion of the results obtained, coming back to my research question. I will conclude by stating

the limits of my study, what we know so far, what we could still analyze, and the next steps to take when researching academic language in general and its vocabulary in particular.

Chapter 1

Academic language

1.1 Definitions of academic language

The concept of academic language is not easy to define. In this chapter, we will say more about academic language as a research topic, including its definitions and properties, as well as acquisition and teaching in both first language and second language contexts; moreover, we will look at a comparison between academic Italian and academic English, and finally at previous studies on academic language learners of both L1 and L2.

Academic language is a register used in educational contexts, including schools and universities, that is characterized by certain features concerning grammar, vocabulary and discourse. Apart from being a medium used for communication in academic contexts, i.e., to produce texts or discourses, academic language is also a competence which is not innate and therefore has to be acquired with time. This acquisition process applies to both L1 and L2 speakers and starts at the early stages of life, as explained by Cummins (1979, in Mutema, 2022), whose work represents one of the first milestones for studies on academic language. Cummins conceived the expressions Basic Interpersonal Communication Skills (BICS) and Cognitive Academic Language Proficiency (CALP), which apply equally to L1 and L2 speakers. With BICS he referred to the basic stage of language learning, which is mainly focused on achieving oral communicative competence. BICS is normally used in conversations in informal contexts, such as home or communities, and the vocabulary associated with this phase is composed of common high-frequency words. By age five, most L1 speakers reach a high level of BICS (Cummins, 2009 in Mutema, 2022), and at that age children usually start going to school, where they meet with the language typical of educational settings, which corresponds to CALP, also referable as academic language. With CALP, students gain competence in tasks like reading and writing, develop critical thinking in various disciplines and start working independently as students (Millin, 2016 in Mutema, 2022). It has been shown that a well-developed BICS in the L1 has a great influence on successful L2 learning; therefore, learners should be encouraged to continue developing their L1s as they continue their studies (Mutema, 2022).

Despite its importance, Cummins' approach is not the only one. Other researchers, such as Bayley, kept away from defining academic language, but focused instead on the skills connected to mastering such a register, as reported by Boscolo and Zuin (2015: 17-21):

“Bayley (2007) ha evitato la difficoltà di definire il termine focalizzandosi invece sulle competenze dell’individuo accademicamente alfabetizzato, il quale conosce e sa usare un vocabolario generale e specifico, strutture grammaticali complesse, funzioni linguistiche e strutture discorsive, allo scopo di acquisire nuove conoscenze e abilità, interagire su un argomento o impartire informazioni ad altri.”¹

The researchers who instead tried to define the concept of academic language have done it in different ways. Schleppegrell and O'Hallaron (2011), for example, define it as “the disciplinary registers that students encounter in the secondary years”, while Zwiers (2008:20, in Mutema, 2022: 6) describes it as “the set of words, grammar, and organizational strategies used to describe complex ideas, higher-order thinking processes, and abstract concepts”, and Flynt and Brozo see it as the “word knowledge that makes it possible for students to engage with, produce, and talk about texts that are valued in school” (2008: 500, in Sekhar Rao, 2022: 2). As remarked by Snow and Uccelli (2009), the expression “academic language” is often substituted by other phrasings, such as “the language of education”, “the language of school”, “the language of schooling”, “the language that reflects schooling”, “advanced literacy”, “scientific language”; this choice of terms reflects the specific context this register refers to. As the expression *per se*, there is also an ambiguity relating to the adjective “academic”, as explained by Mastrantonio (2021), who states that its meaning goes from being associated with university to being related to schooling in general, the latter being influenced by the English meaning of the term. The online version of the Italian dictionary Treccani gives as first meaning of the word the reference to Plato’s Academy, while the second meaning is “Di un’*accademia* in genere, antica o moderna” and the third “Universitario, che concerne l’università o l’insegnamento universitario”². The setting of secondary school is, therefore, excluded from this scenario; this is not the case of the definition provided in English by the online version of Merriam-Webster dictionary³, which is instead “of, relating to, or associated with an academy or school especially of higher learning”. In addition, academic register concerns both written and oral genres, the first including, for example, summaries and essays, while the second includes, for instance, presentations and class debates.

In this thesis, I will focus on academic language as the register used by Italian L1 high-school students for written works.

¹ “Bayley (2007) avoided the difficulty of defining the term by focusing instead on the academically literate individual’s skills, who knows a generic and specific vocabulary, and is able to use it, together with complex grammatical structures, linguistic functions and discourse structures, with the aim to gain new knowledge and skills, interact on a topic or give information to others.” (Translation by A.P.).

² Respectively “Of academia in general, old or modern” and “Relating to university or university teaching” (Translation by A.P.).

³ Definition available at the following link: <https://www.merriam-webster.com/dictionary/academic> (Last visited: 06-11-2023).

1.2 Main characteristics of academic language

Difficulties are found not only when it comes to defining the concept of academic language, but also to describing it. In “The challenge of academic language” by Snow and Uccelli (2009), the authors claim that in the literature it is possible to find many lists presenting the various traits of academic language(s), and while it is clear that these aspects differ a lot from those typical of colloquial language, it is not clear whether they are always present. Despite these difficulties in describing it, as Meiners-Lovel (2020: 85) claimed, “academic language exists and that it has features that can be learned and taught”. In order to be able to present the empirical study I worked on, based on L1 Italian written production of academic language, I provide an overview of some of these features, which I believe to be necessary to comprehend the competences that students need to acquire.

Looking at early stages of schooling, Christie and Derewianka (2008, in Kong and Hoare, 2012) identified the following traits of academic language after conducting an analysis of writing by students attending primary and secondary schools in Australia, focusing on texts considered “good” by teachers in various subjects, such as English, history and science:

1. Use of subject-specific vocabulary;
2. Use of nominalizations;
3. Use of grammatical metaphors⁴;
4. Use of complex noun phrases;
5. High lexical density⁵.

Concerning English and German, Morek and Heller (2012, in Marx et al., 2017), on the other hand, made a summary of the taxonomies redacted by Feilke (2012), Gogolin and Lange (2011), Uesseler et al. (2013) and Vollmer and Thürmann (2010), which can be seen in Table 1.1.

⁴ According to Halliday (1985: 321 in Romero and Soria, 2005: 4), grammatical metaphors are “an incongruent realization of a given semantic configuration in the lexicogrammar”; one method to obtain such a structure is to apply a variation in structure and or grammar classes of the terms involved. Romero and Soria (2005) provide the following example: “The cast acted brilliantly so the audience applauded for a long time” becomes “The cast’s brilliant acting drew lengthy applause from the audience”.

⁵ As explained by Johansson (2008: 65), lexical density refers to “the proportion of content words (nouns, verbs, adjectives, and often also adverbs) to the total number of words”.

Lexical features	Syntactic features	Discursive features
Quality of lexis: diverse, subject specific; Prefix verbs (e.g., <i>to reverse, to preempt, to substitute</i>); Nominal compounds (e.g., <i>bar graph, two-digit number, bottom line</i>); Standardized technical terms (e.g., <i>rectangular, rule of three</i>)	Sentences instead of prosodic segmentation; Local coherence by cohesion markers (e.g., conjunctions), complex sentences (e.g., relative, conjunctive, and disjunctive clauses; infinitival, participle clauses); Mode of representation: declarative mood, impersonal expressions (e.g., agentless passives)	Speaker roles and turn taking organization (pre)determined
Lexical density; Content words instead of pronouns; Nominalizations and elaborate noun phrases (e.g., <i>legalization, editing, average breath-holding capacity</i>)		Monological forms (e.g., lecture, presentation, essay)
		Subject-specific text types (e.g., minutes, report)
		Stylistic standards (e.g., objectivity, well structured, adequate length of text)

Table 1.1: The summary of the characteristics of academic language by Morek and Heller (2012, in Marx et al., 2017).

As we can see, these features are very similar to the ones listed by Christie and Derewianka (2008, in Hoare and Kong, 2012): they cover many different linguistic categories, including grammar, syntax, and vocabulary, and this shows the complexity of academic language. This is the reason why in the literature there is often a comparison between academic language and a more colloquial language,

which is seen as less complex: the context and the situations at home and at school are, indeed, very different. At home or in other informal contexts, children and young people can feel freer to express themselves, and do so without too much effort, often because there is no need to be explicit, since the topics they talk about may be already familiar to their speakers. But at school, students are required to prove their abilities in using a more complex language. Ideally, at the end of their compulsory education, the students should have reached such a level of proficiency that they could easily continue their studies at university; as explained by Gadda (1995: 2; in Schleppegrell, 2004), students are expected

to provide reasoned, concrete, and developed presentations of their points of view [and demonstrate the] ability to control a range of vocabulary appropriate for beginning college students, to manage varied syntax accurately and appropriately, and to observe the conventions of standard written English.

In the same study, Schleppegrell (2004) also remarks the presence of nominalization and high lexical density in texts written for school as opposed to spoken interaction in more informal contexts, while Boscolo and Zuin (2015: 17-21) remark another difference between these two contexts, claiming that academic language is more concise and denser, “nel senso che usa termini più precisi e più parole di contenuto, quali nomi e verbi, ma [appare] anche più organizzato grazie all’uso di connettivi”⁶.

Given the fact that academic language is used in more formal contexts and requires precision and control, we can begin to imagine the difficulties faced by students when learning how to use such a register. Boscolo and Zuin (2015) highlight that when using academic language at school, pupils need to write by putting themselves in the reader’s shoes: this means that not only they have to express themselves by focusing on the kind of language they use, but they also have to constantly keep in mind the target reader for the texts they are redacting. But who is the target audience of students’ writing? Snow and Uccelli (2009, in Boscolo and Zuin, 2015) believe that the difficulty of such writing lies in the fact that the target audience is not clearly defined: the texts written by students will only be read by teachers, even though these works were not specifically aimed at them.

Another list of features of academic vocabulary worth mentioning is the one by Mastrantonio (2021), who mentions interesting phenomena related to writing such as impersonality and the tendency to remove the mention of the agent; these choices are linked with the use of passive, already introduced above, and with the focus brought more on the result of an action than on the person who made it (example: ‘this theory illustrates’; Desideri, 2011 in Mastrantonio, 2021). Despite this, academic texts

⁶ “in the sense that it uses more precise terms and more content words, such as nouns and verbs, but also in the sense that it [appears] more organized thanks to the use of connectives”. (Translation by A.P.).

are full of verbs that create a sort of “dialogue”, as explained by the author: it is easy to find quotes, so naturally there are often verbs such as ‘affirm’, ‘mention’, ‘inform’, ‘say’, ‘reply’ and so on. Other typical constructions are built with modal verbs, for example ‘it can be said’, or can be defined as deontic, such as ‘it is necessary to’, ‘we must’; all these examples come from Desideri (2011, in Mastrantonio, 2021). The graphic and structural parts are important, too, and this is linked with the fact that some of these texts, depending on the genre, are divided in sections, with titles and white spaces and include notes, graphs, tables, indexes, and references (if they are dissertations, for instance); the division in parts is accompanied by expressions such as ‘in the first place’, ‘furthermore’ and ‘finally’. References to the textual deixis are also copious, as shown also by this dissertation: ‘in the previous chapter’, ‘as we will see later’, etc. (cf. Desideri, 2011 in Mastrantonio, 2021)⁷. Another characteristic is the presence of the logical and argumentative relations, expressed through the use of the already mentioned complex structures; one example is the cause-effect relationship, which can be communicated thanks to the use of connectors such as ‘consequently’ and ‘therefore’, as mentioned by Natividad (1999).

As for the features that belong to oral communication, Mastrantonio (2021) mentioned “Le forme della comunicazione accademica: ricerche linguistiche sulla didattica universitaria in ambito umanistico” by Ciliberti and Anderson (1999), focused on characteristics that are typical of this context, such as informative and interpersonal aspects of digressions, succession of speech turns during seminars and co-construction of knowledge between teachers and students in oral examinations.

We have now looked at many of academic language features and we are beginning to understand its complexity, given the fact that it concerns many linguistic categories: one of the most interesting ones, i.e., vocabulary, will be the topic of the next chapter.

1.3 Acquisition and teaching of academic language for L1 and L2

As anticipated above, academic language needs to be learned both by L1 and L2 students.

In the previous section, we have seen the complexity of academic language; its variety makes it almost look like an entire new language to learn, as stated by Zwiers (2014). With Cummins’ approach, we have seen that the mastering of academic language takes time, regardless of the speakers’ native/non-native status; this view is supported also by Ferguson et al (2011: 42), who believe that “academic literacy (...) is not part of the native speaker’s inheritance: it is acquired rather through lengthy formal

⁷ The original examples are in Italian and were translated by A.P.

education and is far from a universal skill”. Naturally, native speakers have an advantage linguistically speaking, but this does not mean that they do not face difficulties when it comes to academic language, since they still need to acquaint themselves with the conventions and the features of academic language and with new meanings of words. Scarcella (2003, in Zwiers, 2014), for example, explains that English native speakers have to learn not only new terms for new concepts, but also new meanings for words they already knew, while in a survey conducted by Berman and Cheng (2010), native university students’ responses indicated that they had difficulties with both writing and reading. Research, however, has not focused as much on the challenges faced by L1 students when learning the features of academic language, as highlighted by Snow and Uccelli (2009); a possible reason behind this is the common belief that native speakers do not face many difficulties when using their mother tongue. This “privileged” status given by nativeness, however, has been criticized by some researchers, such as Römer and Arbor (2009), who showed that native students lack academic phraseological items and therefore hypothesized that experience is a factor that should be more considered than nativeness when it comes to academic language proficiency. Experience is strictly connected to teaching and its importance for L1 pupils has been highlighted also by Ligoš (2020: 6), who stated that L1 teaching mainly revolves around “language and language communication, whose purpose is to enable the pupils to acquire communicative and cultural competences during the process of the integral and comprehensive development of their personalities”.

As for L2 contexts, there are more studies on its relationship with academic language, also because of the important role played by English as a *lingua franca* in academia. It is clear that L2 learners have to face numerous difficulties, since concepts and definitions related to language are newer, and they have to learn them in a new language, too. A teacher interviewed by Barnes (2012) stated that one effective strategy for helping L2 students to acquire academic vocabulary is to first let them write in their mother tongue, and later make them work on the transition from L1 to the L2; sensitivity to cultural elements, however, may hinder the academic language acquisition process, for example if teachers do not provide adequate feedback to students making mistakes by focusing on respecting their native language and culture (Zwiers, 2014).

In general, the acquisition of academic language might be enhanced through some methods and techniques, listed for instance by Zwiers (2014); these include making students participate to debates, where they would have to argue and express their thoughts, making them use academic language to do so; requiring a more elaborated and complex vocabulary from them, also with rephrasing, for example. Students should also be encouraged to read more: it has been proved that those who read more deal better with academic language (Krashen and Brown, 2007). As we have seen, learning academic language is a crucial skill that every student should acquire; for this reason, it is important

to investigate more on the acquisition and teaching of academic language for native speaker students, too.

1.4 Academic Italian vs. academic English

Academic Italian has not been given much attention in the literature; the notion of “academic discourse” first appeared in Italy only at the end of the 1990s (Fiorentino, 2015 in Della Putta and Pugliese, 2020) and extensive discussions about this topic did not start before a decade later (Desideri and Tessuto, 2011 in Della Putta and Pugliese, 2020). According to Mastrantonio (2021:350), academic Italian has been treated in research mainly for didactics:

“L’italiano accademico è stato affrontato perlopiù secondo un approccio didattico: è stata l’esigenza di agevolare la comprensione della lingua per lo studio a spingere gli studiosi a definirne i tratti, riconoscendo delle regolarità che percorrono i testi trasversalmente rispetto ai settori disciplinari.”⁸

In this contest, there is not much consideration of the use of academic language as a communication tool in oral exchanges, but mostly in written production. The use of academic language in writing is also related to the use of specialized languages, since considerations related to such languages can often apply to academic language as well: this register is indeed used to communicate highly-specialized contents of many disciplines (Mastrantonio, 2021). Furthermore, these studies mainly concern university students (Pugliese and Della Putta, 2020): this goes back to the ambiguity of the term ‘academic’, described above, which in Italian seems to not consider the world of schooling in general and refers to the university context only. Moreover, the interest for academic Italian has been mainly linked with research that deals with the difficulties faced by people learning Italian as an L2, and not much on those encountered by native speakers, as stated by Boscolo and Zuin (2015): there are indeed not many studies on Italian as L1, and these few studies have been mainly focusing on the teaching of writing to L1 pupils, in order to understand their most common errors to then develop teaching methods to enhance students’ writing skills (Mastrantonio, 2021); other communication skills such as the ones relating to reading or speaking, for example, have not been examined in depth. While looking at the literature on academic language, it is almost immediately clear that academic English as a research topic is much more explored than academic Italian, both for L1 and L2; the expression ‘academic English’ is indeed often used as a synonym for ‘academic language’ in general:

⁸ “Academic Italian was mostly faced with an educational approach: it was the necessity to facilitate language comprehension for studying purposes that encouraged researchers to define its traits, highlighting some regularities that indirectly cover the texts across disciplinary fields.” (Translation by the A.P.).

this fact is not surprising, given that English has become the *lingua franca* for publications on hard sciences, and is becoming increasingly popular also for the so-called “softer sciences”, including for instance social sciences and communication (Narvaez-Berthelemot and Russel, 2001). English plays a huge role in academia also because of its hegemony in the EU, where in 2020 was the most commonly studied foreign language in the upper secondary schools (96%), according to Eurostat⁹, and because of the high number of degrees taught in English outside UK, the US, Australia and Canada, which were nearly one in five according to a 2021 survey by the British Council and Studyportals¹⁰.

1.5 Previous corpus-based studies on learners’ academic language

Many studies on academic language are based on lists of the most important words, which we will analyze in detail in the next chapter; however, some have corpora as a basis without considering word lists. A corpus is “a systematic collection of texts, which can be written or spoken, and can be stored on a computer” (Casañ Pitarch, 2016: 4, referring to Fillmore, 1992 and O’Keefe et al., 2007). The branch of linguistics dedicated to corpora is called corpus linguistics. According to Bennett (2010), studies of languages through corpora first started around 1910, even though the first computer-based corpus, the Brown Corpus, which contains about one million words, was created only in 1961 by W. Nelson Francis and Henry Kučera (Brown University, USA).

An interesting type of corpus for studies on academic language is the ‘learner corpus’; Granger (2008) gave a definition of this type of resource, by stating that

“learner corpora are electronic collections of language data produced by L2 learners, that is, second or foreign-language learners. This relatively new resource is of great relevance for both second language acquisition (SLA) research and foreign-language teaching (FLT)”.

Learner corpora provide a useful basis to analyze and understand the way students use academic language, but unfortunately, the availability of corpora of students writing is historically limited (Nesi and Gardner, 2012 in Durrant, 2016). A very comprehensive list of learner corpora is regularly updated by the Centre for English Corpus Linguistics (CECL)¹¹, part of the Institute for Language

⁹ Article available at the following link: : <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html> (Last visited: 07-11-2023).

¹⁰ Resource available at the following link: <https://www.porta.eurac.edu/> (Last visited: 07-11-2023).

¹¹ List available at the following link: <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html> (Last visited: 07-11-2023).

and Communication (ILC) at the Université catholique de Louvain, and multiple studies carried out having learner corpora as a basis can be found also on the PORTA website¹², which also provides the corpora related to these studies and created by the Institute for Applied Linguistics at Eurac Research (Bolzano, Italy).; in most cases, the term ‘learner corpus’ refers to L2 contexts.

As we can imagine, there are many studies on the use of various language features’ use in academic contexts. According to Biber, research on this topic became more popular as linguists started to think it was important to teach “the specific kinds of a language that a learner will need” (2006: 6). In order to prepare adequate teaching material, Biber (2006) suggests that it is necessary to first identify the linguistic features of academic register; a cooperation between the world of research and teaching in schools might be useful indeed in order to help students facing difficulties in education contexts, and he also states that many studies focused on academic language are indeed being published. In this section, I will give a review of some corpus-based studies focused on L1 and L2 academic language learners that contributed to the research on this topic.

1.5.1 Previous studies on L1 learners’ academic language

A study based on an English L1 corpus of student writing is the one conducted by Hardy and Römer (2013), who analyzed a part of the Michigan Corpus of Upper-level Student Papers (MICUSP, 2009), developed by researchers at the English Language Institute of the University of Michigan and composed of texts written by upper-level students about different disciplines that were graded with A. The authors of the analysis used this corpus to identify lexico-grammatical features that might be helpful to describe successful student writing. As we have seen, writing is perhaps the communication skill that is most connected to academic language and is surely one of the most studied in relation to this register.

An innovative study based on an Italian L1 corpus is ITACA, resulting from the project “ITACA - Coerenza nell’ITALiano Accademico” (Bienati et al., 2022) conducted by a team of the Institute for Applied Linguistics at Eurac Research (Bolzano); the main aim of this research is to measure the level of textual coherence in a collection of argumentative texts written by L1 Italian high-school students. This corpus is the one used in this thesis for the analysis of academic vocabulary, which will be presented in the last chapter.

Concerning the comparison between L1 and L2, Della Putta and Pugliese (2020) conducted a study on written academic discourse in exams taken by university students with Italian as L1 and L2. For

¹² Resource available at the following link: <https://www.porta.eurac.edu/> (Last visited: 07-11-2023).

their research, they used a corpus made of 58 texts written for the examination of the course “Teaching L2 Italian” offered by the M.A. in Italian culture and language for foreigners at the University of Bologna: 40 texts were written by Italian native speakers, while the remaining 18 by foreign Erasmus students who have Italian as L2. This corpus was collected with the aim of analyzing both the similarities and differences in academic writing between two groups of students with a different level of knowledge of Italian, and the factors on which the research was focused included text length, vocabulary choices, text organization, textual coherence, and readability. After comparing the texts written by the two groups, the authors noticed that, while L2 speakers are capable of using specialized vocabulary and of structuring the text, the L1 speakers’ texts lacked coherence and contained language characteristics that are typical of oral communication¹³.

1.5.2 Previous studies on L2 learners’ academic language

A study focused on the comparison between the production of L2 students and more experienced speakers was the one by Liu and Li (2016), who analyzed in particular the complexity of noun phrases by using two different corpora; the first contains 70 dissertations written by Chinese postgraduate students with English as a foreign language (EFL), while the second is composed of 129 empirical research articles, written by people with different L1 backgrounds, taken from six journals in applied linguistics. The authors explain that most of the research focusing on academic texts privileges aspects concerning structure and citations, ignoring students’ issues with lexis and grammar, which are relevant since many students in academia have to write in English as L2. The student corpus showed lower noun phrase complexity, and according to the authors this result can be used “as a practical guide for raising students’ awareness of the degree of complexity of multiple postmodification common in published academic writing” (2016: 62)”. Multiple postmodification is defined by the authors themselves as the process of making a noun phrase complex in a very elaborated way¹⁴. They also state that students should be encouraged to compare what they write with what experts write instead; being more acquainted with adequate examples of this type of writing could be useful for students, who as we have seen face many difficulties when dealing with academic language; a comparison of this kind could help them understand better what they should aim for with their work.

¹³ A possible explanation for this phenomenon lies, according to the authors, in the differences in the educational system of different countries: in Italy, students are not quite used to written evaluation tests, as opposed to foreign students (Fallani and Troncarelli, 2018, in Della Putta and Pugliese, 2020); however, the foreign students’ origin is not mentioned in the study.

¹⁴ An example provided by Li and Liu (2016: 54) is the following: “an understanding **of** language **as** an ideational signifying system **that** plays a central role **in** how we understand ourselves and the world”.

As for L2 Italian, Corino and Marelo (2009a) used the texts from the VALICO corpus (Varietà di Apprendimento della Lingua Italiana Corpus Online; Corino and Marelo, 2009b¹⁵), composed of texts written by L2 Italian learners, for creating exercises based on a multiple choice task, including also the errors made by learners. These texts, according to the authors, paved the way for reflections on grammar, highlighting the usefulness of learner corpora. One of the main studies related to academic language and aimed at improving the learners' academic language knowledge was conducted by Stefania Spina, who created the Academic Italian Corpus (AIC) in 2008. This corpus is a collection of 240 texts that cover three discipline groups (humanities, law and economics, science) and encompass four different genres, all pertinent to academia (handbooks, lecture notes, scientific articles and dissertations); from this corpus Spina extracted a word list of Italian academic language, which was used for the present study and will be described in the next chapter.

By looking at this overview and at the state-of-the-art of studies on academic language both for L1 and L2 students, we can conclude that the features which have been identified as typical of this register can be analyzed and extracted thanks to research resources such as corpora, and can then be taught to students, who can use them to express ideas and communicate contents on numerous topics in education contexts, following the specific conventions of academic language. However, we have also noticed that research on L2 is much more present than that on L1, given the importance of English as L2 in academia; for this reason, studies on how students deal with academic language in their L1 should be more valued and supported.

¹⁵ “Variety Learning Italian Language Corpus Online” (Translation by A.P.); more information about the corpus available at the following link: www.valico.org (Last visited: 17-11-2023).

Chapter 2

Vocabulary and word lists

2.1 Definitions and types of vocabulary

The term “vocabulary” refers to all the words in a language or that are known by an individual. Vocabulary teaching has been an important study object for some time now (see e.g., Laufer and Nation, 2012). While analyzing vocabulary, we can see that the notion of frequency is almost always present; this can be explained by the fact that despite studies focused on vocabulary saw a growth in the 1990s, lists of the most needed vocabulary have a longer history (Laufer and Nation, 2012) and have often been based on the criterion of frequency: the characteristics related to this criterion often appear in vocabulary research, and types of vocabulary are defined accordingly.

Schmitt and Schmitt (2014), for example, divide English vocabulary into the following groups:

- High-frequency vocabulary
- Mid-frequency vocabulary
- Low-frequency vocabulary

‘High-frequency vocabulary’ is composed of 3,000 word families. As explained by the authors, a word family consists of a root form, its inflections, and its derivatives; an example provided by them revolves around ‘select’ as root form: its inflectional forms would be ‘selected’, ‘selecting’, ‘selects’ and its derivatives ‘selection’, ‘selective’, ‘selectively’ and ‘preselect’. Traditionally, this first section was composed of 2,000 word families; however, this boundary has never been established in a precise way, despite being often mentioned in teacher guidebooks and research publications, as stated by Schmitt and Schmitt, who also believe this threshold to be influenced by the General Service List (GSL; West, 1953), a word list composed of 2,000 lemmas believed to be the most frequent and therefore important to learn in English. Given that much research has been done on vocabulary since GSL’s publication, the authors decided to see if an adjusted figure would prove more useful nowadays. After analyzing various research studies based on samples of both written and oral communications and on the number of words included in the defining vocabulary lists of English learner dictionaries, which are composed of those lemmas that the publishers consider to be the most useful for covering a great variety of meanings, the authors came to the conclusion that 3,000 seemed a more appropriate threshold for this type of vocabulary.

‘Mid-frequency vocabulary’ covers the range between the 9,000 and the 3,000 levels. This category had not been clearly defined before Schmitt and Schmitt’s study, and they also indicate some of the benefits of knowing words from this set, deriving from the capacity of dealing with language for authentic purposes like watching movies and reading textbooks; another positive impact of the knowledge of this vocabulary group was also seen on the improvements in reading test scores and in general on the higher fluency of students.

Finally, the boundary of ‘low-frequency vocabulary’ is set at 9,000 word families, against the traditional 10,000; this decision was made after a corpus study conducted by Nation (2006), who analyzed a series of English authentic texts and calculated that 8,000-9,000 word families were enough to reach a coverage of 98%, which is the percentage believed to enable efficient reading. Schmitt and Schmitt then used Nation’s frequency lists on the Corpus of Contemporary American English (COCA; Davies, 2008), which is composed of more than 400 million words and is believed to be the most important corpus of general English (Schmitt and Schmitt, 2014), and saw that the 9,000 most frequent word families covered 95.5% of the COCA; therefore, this new threshold has been considered more accurate than the older one.

In “Learning Vocabulary in Another Language” (2001), Nation identified the following four groups of vocabulary instead:

- High-frequency vocabulary
- Academic vocabulary
- Technical vocabulary
- Low-frequency vocabulary

For Nation, ‘high-frequency vocabulary’ includes not only function words such as prepositions or articles, but also many content words, i.e., nouns, verbs, adjectives, and adverbs. This level is also set at 2,000 words, as anticipated before, and it corresponds with West’s GSL, which covers 80% of the words in academic texts and newspapers, and around 90% in conversations and novels.

This list was also the basis for the Academic Word List (AWL; Coxhead, 1998), which is in turn the most commonly used list of academic words and is made of the 570 words that make up the ‘academic vocabulary’ according to Nation. ‘Technical vocabulary’ consists instead of those words that are frequent in a specialized text or in a specific field, but not elsewhere; they are therefore defining of the domain on which the focus is. They usually cover 5% of the total words in a text, and the same percentage goes for ‘low-frequency vocabulary’, the last and biggest group identified by Nation,

which is made up of the words that did not make it to the other groups, e.g., proper nouns or rarely used terms.

The attempts to divide vocabulary in bands are strictly related to the searching for the number of words needed by individuals: given the differences among languages there is no estimation of this value that can be applied in all cases; however, it is believed that one needs to know approximately 95% of a text in order to understand it (Nation and Waring, 1997).

To conclude, after looking at these categorizations of vocabulary, we can say that it is not easy to strictly divide a linguistic category characterized by variety, since the lines between the different groups that were just described might sometimes be not so well-defined.

2.2 Vocabulary knowledge and acquisition

We have just seen that research has been focusing on how many words one needs to know, but what does it mean to know a word? This is another crucial question on which research has long been focusing. In this section, I will give an overview of two important processes related to vocabulary, i.e., its knowledge and its acquisition.

One of the first lists of elements making up word knowledge was made by Cronbach (1942, in Milton and Fitzpatrick, 2017), who believed that this knowledge was made from the following concepts: generalization (the word's definition), application (proper use of a word), breadth of meaning (a word may have more than one meaning; this is a reminder to the concept of polysemy), precision of meaning (the capacity to use the different meanings in a correct way in various contexts), and availability (the productive use of a word, i.e., its correct use, different from the ability to just recognize a word, which is known instead as receptive use).

The main work on vocabulary knowledge is probably the one by Richards (1976, in Meara, 1996), who created the following list, which includes eight assumptions about word knowledge (1976: 83):

1. The native speaker language continues to expand his vocabulary in adulthood, whereas there is comparatively little development of syntax in adult life.
2. Knowing a word means knowing the degree of probability of encountering that word in speech or print. For many words, we also know the sort of words most likely to be found associated with the word.
3. Knowing a word implies knowing the limitations imposed on the use of the word according to variations of function and situation.
4. Knowing a word means knowing the syntactic behaviour associated with that word.

5. Knowing a word entails knowledge of the underlying form of word and the derivatives that can be made from it.
6. Knowing a word entails knowledge of the network of associations between that word and the other words in language (sic.)
7. Knowing a word means knowing the semantic value of the word.
8. Knowing a word means knowing many of the different meanings associated with the word

As hinted by Milton and Fitzpatrick (2017), on the one hand, this list repeats concepts that had already been expressed in the past, for instance by Cronbach's list (1942), like points 2, 3, 7 and 8; on the other hand, it also introduces some interesting points relating not just to words in isolation, but also to lemmas in connection with each other, e.g., points 4, 5, and 6. However, these points should not be analyzed by themselves, but rather be seen as steps of a learning process.

The number of known words is called vocabulary size or breadth, while the level of knowledge of words is referred to as vocabulary depth, as explained by Webb (2013). Since this research does not focus on how well words are known, I will examine in more detail the concept of vocabulary breadth. There are many researchers who tried to create tests aimed at measuring vocabulary size; one example is the Peabody Picture Vocabulary Test (Dunn, 1959), during which the examinee is required to choose the picture (from a sequence of four) that they consider the best to represent the meaning of a word. Other measures of vocabulary breadth include the Yes/No tests, the firsts ones of which were used for L1 students (see e.g., Anderson and Freebody, 1982), who had to simply indicate whether they knew a word or not in a list including also "nonwords", i.e., words that look like they carry a meaning but actually do not, such as *porfume*; the Vocabulary Levels Test (VLT; Nation, 1983, updated by Schmitt et al., 2001), where students of English as L2 are asked to link words to definitions; the Vocabulary Size Test (VSL; Nation and Beglar, 2007 and updated by Cohead et al., 2014), based on a multiple choice of definitions for a word and used for students with English as both L1 and L2.

According to Wilson et al. (2016), vocabulary knowledge is at the heart of diagnostic tests for language and educational acquisition in general and plays an important role in predicting both linguistic and cognitive abilities from the first stages of life; this means that vocabulary knowledge provides a base for the development of communication skills, which are also used in academic contexts. The concept of vocabulary knowledge is strictly connected with the process leading to it, i.e., vocabulary acquisition, which has a level of consciousness or unconsciousness; being exposed to inputs is sometimes not intentional, and one could wonder if vocabulary teaching at school should be limited to providing inputs or if it should (also) be based on explicit instructions. We have seen the

importance of vocabulary knowledge and acquisition and their relationship to communication skills which are used also in academia; for this reason, it is important to continue to investigate on them and how they can be enhanced and developed, for both L1 and L2 students; the relationship between vocabulary knowledge and acquisition in L1 and L2 will be examined in depth in the following subsections.

2.2.1 Vocabulary and L1

Of course, L1 speakers have a higher level of vocabulary knowledge compared to L2 learners, and vocabulary knowledge is connected to communication skills in general, as we have mentioned before; a demonstration of this is a study conducted by Karalık and Merç (2019), who concluded that there is a positive correlation between L1 vocabulary knowledge and students' listening comprehension. As for the acquisition of vocabulary in L1, Alharbi (2019) believes that it "functions in a systematic way, as semantic features are developed according to the learner's time exposed to the language". L1 vocabulary acquisition has not been a popular research topic, despite its importance: Nation and Waring (1997) believe indeed that analyzing native speakers' vocabulary size could be useful also when working on vocabulary teaching for L2 learners. Research on L1 vocabulary has been mainly focusing on intentional learning, even if as children we learn most of our native vocabulary through incidental learning, for example by reading, as stated by Hulme (2018); in her study, she explains that the comparison between these two types of L1 vocabulary acquisition find little space in the literature, but it should be important to start focusing more on them, since different approaches might lead to different results in acquisition. Her experiments, based on the reading of stories to pupils, allowed her to conclude that "new word meanings are learned more efficiently under intentional learning conditions than incidentally through story reading. However, there was also some evidence of less forgetting of items learned through stories, suggesting that word meanings learned in a more semantically rich context could be retained better" (2008: 66). Learning through stories was the focus chosen by Elley (1989), too: in this study, teachers in New Zealand read stories to 335 L1 pupils aged 7 and 8 in New Zealand. After an analysis on the acquisition of the words that were mentioned in these stories, it emerged that children incidentally learned many lemmas, and subsequent tests proved that this learning was not temporary; moreover, the presence or absence of an explanation for the word meaning given by the teachers was not relevant. Incidental vocabulary learning has then proved to be successful, as remarked by Meganathan et al. (2019: 53), who stated that it "endorses deeper mental processing and better recall".

As for Italian specifically, according to D'Aguanno (2019) vocabulary teaching does not play an important role in traditional Italian teaching since it is often limited to the presence of notes next to texts, where the less known words are explained, and to the corrections by teachers in both written and oral tests. He also suggests a solution for choosing which words should be taught to students: apart from looking at the specific needs of the young pupils, which vary also inside one classroom, it could be useful to respect three criteria: the general frequency of vocabulary, the context for its use, and the communicative function in place. Concerning academic vocabulary in particular, D'Arguanno (2019) states that it is necessary both to read school texts and to write those texts, as high schools should teach to do; thanks to academic vocabulary, students can elaborate their thoughts and will see the cognitive development connected to writing enhanced. As for English, Duffelmeyer (1980, in Ortega, 2017) conducted a study to see if the teaching L1 English vocabulary based mainly on experience led to better outcomes compared to traditional school techniques; a vocabulary test was carried out by 56 U.S. undergraduates, and the results were "in favor of the experimental vocabulary learning procedure enhancing later recollection and application of words in context" (Ortega, 2017: 56).

2.2.2 Vocabulary and L2

Nation (1983) states that vocabulary size is relevant also for EFL learners; in particular, it has been shown that a low vocabulary knowledge might cause stress during the learning process and students may therefore feel demotivated (Al Zahrani and Chaudhary, 2022), while it has been proved that a higher L2 vocabulary knowledge has a positive impact on the improvement of reading proficiency (Tong et al., 2023), and on listening comprehension (Lange and Matthews, 2020).

In general, L2 learners might have a lower level of vocabulary knowledge compared to native speakers, but their acquisition process is different and somehow simplified, since they already possess the conceptual meaning of words in their L1, so new lexical forms in their L2 are acquired into them, as explained by Pavičić Takač (2008, in Alharbi, 2019); a method to enhance L2 vocabulary acquisition that is considered effective is indeed the use of mnemonic devices, which may be visual or not, where the new L2 term is associated with a semantically related L1 word (Hulstijn, 2003, in Borawski, 2019); another technique proved to be successful is extensive reading, according to a study by Kalogeros and al. (2023) conducted on Spanish EFL learners aged 9 and 10.

Research on L2 vocabulary is certainly extensive; it is important to highlight once more the necessity to expand also our knowledge of the processes faced by L1 students, in order to comprehend better

the ways in which words are learned by them, which are connected also to the learning of language in general, including registers such as academic language.

2.3 Word Lists

As anticipated, one of the major questions when it comes to vocabulary is how many words one needs to know; another important question is what kind of words are needed specifically. A resource strictly related to both vocabulary knowledge and acquisition that was created for vocabulary teaching and learning are word lists, which were initially aimed at L2 students, also to help them learn the vocabulary of a new language better. This presupposes that, even if students meet a lot of different words during their studies, most texts contain a relatively small quantity of frequent words, as stated by Nation and Waring (1997). These lists are often based on the criterion of frequency of use in the target language, usually in a corpus, and frequency is generally paired with data about the number and types of texts in which these words appear and about the distribution across different genres, too. As we have mentioned above, if language is made of a relatively small number of words with high frequency, making a list of these words so that learners can focus on them can be seen as an effective strategy when learning a second language, and word lists could then be used also for defining a lexical minimum for a certain language, i.e., for selecting words to teach at a specific proficiency level. In the next subsections, we will describe some of the best-known word lists for English and Italian.

2.3.1 English word lists

The most used English word lists are Michael West's General Service List (GSL) and the Academic Word List (AWL) by Averil Coxhead.

The GSL is a general-purpose list which contains the most frequent words in English and was created in 1953 and is a reissue of a report by Faucett and al. (1936) on English vocabulary selection that was published after two international conferences on the role of word lists for EFL in education (Gilner, 2011). This list counts 2,000 headwords, which are then further divided into two lists of 1,000 English words; the lists are frequency-based and each one is in alphabetical order; each headword is part of a word family, e.g., 'you' is the headword of the family 'your', 'yours', 'yourself', but there are some errors in these groupings, for example 'efficient' was included in the word family of the headword 'effect'. Despite these mistakes and its age, the GSL is still considered one of the best word lists given the information it provides about the frequency of the words' different meanings.

The AWL goes back to year 2000 and is used primarily in English for academic purposes (EAP) classrooms, in many vocabulary tests and it also serves as a very important resource for researchers (Coxhead, 2011). This list counts 570 words and it was based on a written academic corpus of 3.5 million words, divided into four discipline-based sections: arts, commerce, law and science. The words contained in the GLS were excluded from AWL, as the purpose of the latter is to give a representation of academic vocabulary, not general English¹⁶.

A more recent addition has been Gardner and Davies' Academic Vocabulary List (AVL), developed in 2013, which consists of 3,015 words and was created on the basis of the already mentioned COCA by Mark Davies (see Section 2.1). Unlike AWL, the AVL was defined by using lemmas instead of word families; the two lists differ also in their approach to the exclusion of non-academic words: in AVL, words were selected only if they occurred with a ratio of at least 50% higher in the academic corpus than the non-academic one (Smith, 2021). Furthermore, the authors of AVL found that the first 570 words of their list, comparable to the totality of AWL, cover the 14% of the academic materials of both the COCA and the British National Corpus (BNC) (Gardner and Davies, 2013).

2.3.2 Italian word lists

At the moment, the two main word lists for the Italian language are the “Nuovo vocabolario di base” (Nvdb) by Tullio De Mauro and Stefania Spina’s Academic Italian Word List (AIWL).

The “Nuovo vocabolario di base” was published by De Mauro (2016) as an update of the previous list “Vocabolario di base” (Vdb). The original list was published in 1980 as an annex to De Mauro’s book “Guida all’uso delle parole” and was based on data about frequency, use, and availability; for the first two criteria, data was extracted by a frequency list of written Italian, the “Lessico di frequenza della lingua italiana contemporanea” (LIF), created by Bortolini et al. (1972), which was built from a collection of texts with 500,000 words; for the last criterion, data was retrieved from the results of a survey managed by Elisabetta Bonucci, Stefano Gensini and Emilia Passaponti between 1979 and 1980 and conducted on adults having obtained a certificate of lower secondary education (the school attended by pupils aged 11-13. The Vdb included around 7,050 words, divided in three sections: fundamental vocabulary (“vocabolario fondamentale”, FO; FO refers to *fondamentale*), composed of 2,000 words covering 90% of words in non-specialized text or in a discourse; high-use vocabulary (“vocabolario di alto uso”, AU; AU refers to *alto uso*), composed of 2,750 words covering only around 4% of occurrences, but still being much more frequent than the rest of vocabulary, and understood by

¹⁶ This decision led to some criticism, since in the GSL a lot of high-frequency words pertain to academia, such as *company, interest, business, market, account, and capital*, as pointed out by Smith (2021).

at least 50% of middle school students (consequently aged around 11-13) from various Italian regions; high-availability vocabulary (“vocabolario di alta disponibilità”, AD; AD refers to *alta disponibilità*), which counted around 2,300 words considered to be highly available. In the new version, there are 7,500 words selected according to the same criteria of the previous list; the division in parts has also stayed the same. The first section, FO, counts 2,000 words with high frequency that are used in 86% of texts and discourses; the second one, AU, is made of 3,000 words that are used frequently and cover 6% of occurrences; the third one, AD is composed of 2,500 words that are used only in some contexts, but are understood by all speakers and are characterized by an availability which is equal or even higher compared to the most used words. Nvdb has a set of texts with 18,843,459 occurrences as a basis, ranging from press and essays to literature, entertainment, and media communication. De Mauro’s decision to update his original work was based on the hypothesis that in the years that had passed since the original list’s publication, Italian society had faced many changes that impacted vocabulary, too (De Mauro, 2016); he also explained that “use” refers to the product of the absolute frequency of a word’s occurrences in a sample of texts divided in various categories, multiplied by its dispersion, i.e., the number of text categories in which the word occurs.

The AIWL is the Italian equivalent of Coxhead’s AWL and was created by Spina (2010). Unlike AWL, this list contains both words (403) and collocations (280), retrieved from an analysis based on a written academic corpus, the AIC, which was described in the previous chapter. In this case, too, words belonging to the general list for Italian, the Vdb, were excluded; things changed, however, with the updated version of Vdb, as we will see in the next chapter. AIWL covers about 5% of an academic written text, while Vdb takes up about 78%: together, general and academic vocabulary from the lists represent 83% of the words used in communication in academic contexts (Spina, 2010). As for the collocations, as explained by Spina (2010), the process of finding them started with a manual tagging of a collocations list which had been previously extracted from contemporary Italian corpora; this led to the identification of the most frequent sequences of grammatical categories, and for the academic context five of them were selected: adjective-noun (*netta distinzione*, ‘clear distinction’), noun-adjective (*prospettiva teorica*, ‘theoretical perspective’), noun-preposition-noun (*bagaglio di conoscenze*, ‘wealth of knowledge’), verb-article-noun (*affrontare il tema*, ‘address the topic’), verb-noun (*fare riferimento*, ‘refer’). The result of this procedure was then filtered through the value of the coefficient of use, which had to be higher than 2.

With AIWL, Spina hoped to contribute to an accurate description of academic Italian language in order to offer students with an L1 different from Italian a useful resource for gaining more knowledge about Italian academic vocabulary. Another interesting list is the “Lessico della conoscenza” (LC;

Ferreri, 2005), which contains 255 lemmas used in different disciplinary fields and was created in particular for students attending secondary school or university, as explained by Gallina (2019).

2.4 Studies based on word lists

In the first chapter, we have seen how corpora can provide a great basis for studies on academic language(s), and we anticipated that word lists are a useful tool for research on this topic, too; in this section we will give an overview of some studies focused on word lists that contributed to the research on academic vocabulary.

As for English, Li and Qian (2010) measured AWL presence in a financial corpus with the final aim to find ways to teach AWL lexical items that were present in the corpus, and tested two programs, linked to the word list, that proved to be useful for teaching purposes. The first tool, AWL Highlighter¹⁷, gives the possibility to identify words from AWL in a text; while the second tool, AWL Gapmaker¹⁸, creates exercises with gaps that need to be filled with words from the list. Word lists can then be used also for teaching purposes, as we will describe in the next section. A similar study was conducted by Durrant (2016), who analyzed the presence of words from the AVL in a collection of texts written by university students, extracted from the British Academic Written English (BAWE; Nesi and Gardner, 2013). The mean coverage of the list was 34%, a good result according to the author, who states that word lists are a useful learning investment for students, but also acknowledges differences concerning text disciplines.

As for Italian, Dal Negro (2016) tried to measure passive competence of AIWL in a target group composed of 24 people (students, PhD students, administrative staff) at the University of Zurich and at ETH Zurich, with a different level of knowledge of Italian (from A0 to B1, according to the Common European Framework of Reference for Languages levels), by creating a translation test which was later taken by the target group. The results showed that the participants knew more than 58% of words from AIWL, even if they were not explicitly taught in class. Another study, this time based on Nvdb, was conducted by Forti et al. (2019); the aim of this research was to evaluate how effective an automatic tool was for determining the complexity of a text in the context of Italian as L2, called MALT-IT¹⁹. The results were considered satisfactory by the authors, who used Nvdb for the lexical features used to predict the text level.

¹⁷ Tool available at the following link: <https://www.nottingham.ac.uk/alzsh3/acvocab/awllhighlighter.htm> (Last visited: 18-11-2013).

¹⁸ Tool available at the following link: <https://www.nottingham.ac.uk/alzsh3/acvocab/awlgapmaker.htm> (Last visited: 18-11-2013).

¹⁹ Tool available at the following link: <https://lol.unistrapg.it/malt/index.php> (Last visited: 18-11-2023).

From this brief review we can conclude that word lists are a useful tool when it comes to vocabulary, but they also have some limitations; the use of word lists for teaching and the issues highlighted by some of the studies on them will be examined in depth in the following subsections.

2.4.1 Use of word lists for teaching

According to Lessard-Clouston (2013), word lists can be used in many ways for vocabulary teaching, and they offer possibilities that go beyond to the mere memorization of words; a first suggestion concerns the practical use of the words from the lists in various contexts, for example by making sure that students are often exposed to them not only in their writing tasks, but also in the texts they have to read or in the multimedia contents they are showed; in general, teachers should hopefully know the learners' needs, so they should be able to organize their teaching techniques accordingly. Another strategy could be focusing on a section of word lists, since they include many words and learning them all at once would be overwhelming; students might be encouraged to focus, for example, on the contexts in which words they already know are used, or to look at the ways words can be connected to each other to form new meanings; in this way, lemmas are not acquired by themselves. Looking at these suggestions, we could say that they could be easily followed for the teaching of both L1 and L2 vocabulary, even if word lists were initially created for L2 learners; according to D'Aguanno (2019), the work by De Mauro on Nvdb, for instance, could indeed be a great way to start teaching vocabulary in Italian primary and secondary schools. He believes, for example, that words marked as "fundamental", such as *emergere* ('to emerge') or *esigenza* ('necessity') could be unknown or partially known by Italian L1 students attending the first years of secondary schools, and states that AIWL could also be a useful tool for teaching academic vocabulary in an L1 context in secondary schools, together with Ferreri's LC, which was initially aimed at L1 Italian students (Gallina, 2019). We have seen that word lists can then be used also to improve the vocabulary knowledge of native speakers, too, and that their initial scope could therefore be extended.

2.4.2 Problems with word lists

Despite their usefulness for teaching purposes, word lists have also been facing many criticisms. Concerning academic lists, one of their main weaknesses that we anticipated above is the fact that, given that the words mentioned are not pertinent to a specific domain, it is not possible to adequately represent many different disciplines at the same time. For this reason, some researchers such as Chen

and Ge (2007, in Durrant, 2016) suggested that instead of listing general academic words, it would be better to teach vocabulary according to single fields of study, also because words can have different meanings based on the domain, as explained by Martinez and al. (2009), who mentioned the example of the lemma ‘volume’, whose meaning in science is not the same as the one in social sciences.

Word lists include many words, but do not provide indeed information on their use, even when it would be useful for learners; moreover, most word lists include single words only. To this end, Norbert Schmitt (2014: 914-915) expressed his concerns related to the fact that these factors – lack of information on words and on how they can be connected to each other – might lead to a partial knowledge of words:

Studying the words in isolation without contextual elaboration limits the students to learning only something about the word form, something about the meaning, and some linkage between the form and meaning.

The complexity of knowing a word was explained also by D’Aguanno (2019: 95):

“La conoscenza delle parole non va intesa soltanto come conoscenza dei significati, ma come padronanza dell’intera ‘informazione lessicale’, ovvero come padronanza di tutti gli aspetti relativi alla forma (ortoepia, ortografia e morfologia), alla semantica (significato principale, accezioni e relazioni semantiche) e all’uso (valenze, reggenze, collocazioni e registro) delle parole”²⁰

Collocations, mentioned in the quote above by D’Aguanno, represent indeed a potential issue concerning word lists in general; among the word lists we saw earlier, Spina’s AIWL is the only one to include a separate section dedicated especially to collocations. There are, however, two word lists dedicated to collocations exclusively: the first is the Academic Collocation List (ACL), which contains 2,469 collocations, created by Ackermann and Chen (2013), who based their work on the Pearson International Corpus of Academic English (PICAE; Ackermann and al., 2013) and were assisted by English teachers, in order to choose the most useful collocations for students; the second is the Phrasal Expression List, created by Ron Martinez and Norbert Schmitt (2012), a collection of 505 multiword items, selected mainly according to the criteria of frequency and meaningfulness. Creating a word list for collocations, however, is seen as problematic by Durrant (2009); according to him, in the context of EAP specifically, research has been focusing only on “lexical bundles”,

²⁰ “The knowledge of words does not have to be interpreted just as knowledge of meanings, but as competence of the entire ‘lexical information’, i.e., as competence of all aspects related to form (orthoepy, orthography and morphology), to semantics (first meaning, senses and semantic relations) and to use (values, rections, collocations and register) of words” (Translation by A.P.).

which he defines as “frequently recurring fixed sequences of words” (2009: 158). These constructions are easy to find by using corpus search methods, but the lemmas in a collocation do not always appear right after the other: they can be separated by other words, and in these cases this kind of expressions cannot be retrieved easily²¹.

Another problem exposed by Durrant (2016) concerns word lists composed of word families, such as the AWL; he explains that word families are made by “headwords plus their inflectionally and derivationally-related forms” (2016: 4), but the forms related to a word do not all have related meanings: one example reported by Durrant (2016) is ‘constitute’, which in the AWL is linked also with ‘unconstitutional’, despite the semantic difference between these words (being a part of a whole vs. not conforming to the Constitution).

It seems that, as we could have imagined from the previous chapter, vocabulary is hard to define as well, given its various types and domains of uses, and we have also noticed the complexity of word lists: they are a useful tool for teaching purposes, concerning L1, too, but they also have some limitations, since factors such as disciplines, contexts and multi-word expressions pose many challenges for their effectiveness. In the next chapter, we will examine in depth how word lists are used for corpus-based research: we will analyze a corpus of L1 Italian student writing by using the two main Italian word lists described above, to see if they are representative of the way young pupils write today.

²¹ Some examples provided by Durrant (2009: 158) are the following: “he made a **powerful argument**”; “he made a **powerful**, but ultimately unconvincing, **argument**”; “his **argument** was a **powerful** one”.

Chapter 3

Empirical study

3.1 Research question

As stated in the introduction to this dissertation, the aim of this dissertation is to find out how relevant Italian word lists are for texts written by native speakers, in particular if the words present in the lists I chose for my research are representative of the way young people write today. The results of my analyses, based on two Italian word lists and an L1 Italian corpus of student writing, will allow me to see if word lists can be a useful resource for native speakers in academic contexts. The main focus of the study is represented by the percentage of words from the four reference lists I used that are present in the texts composing the ITACA corpus. Other data include the words that are repeated across the four lists, the first 100 most frequent words and the first 100 most frequent N-grams in the corpus. In this chapter, I will present both the analyses I performed, including the process leading to them, and their results, to then wrap up with a discussion of the findings and some of the possible implications for teaching.

3.2 Method

I had to perform various tasks before obtaining the results I needed for my study. In this section, I will present the steps of my research: first, I will describe the corpus I used and the data included in it and also in the word lists, including a presentation of the tools I used for consulting them; finally, I will present the analyses I performed.

3.2.1 The corpus

As anticipated, the corpus I used for this study is ITACA, which is the result of the ITACA project, conducted by a team of the Institute for Applied Linguistics at Eurac Research (Bolzano)²². It is composed of 636 texts and 387,772 tokens, written by Italian L1 students at their fourth year of high school (“scuola superiore di secondo grado”) in the province of Bolzano, consequently being around

²² A presentation of the project (Bienati et al., 2022) is available at the following link: <https://italianoascuola.unibo.it/article/view/14830/15638> (Last visited: 07-11-2023).

18 years old. The texts have a specific topic: the participants were asked to write a letter imagining it was addressed to the Italian Minister of Education, who has hypothetically decided to permanently implement blended learning even after the end of the sanitary emergency caused by the COVID-19 pandemic. In this letter, students had to express their opinion about this choice, stating pros and cons of this way of learning. The texts were saved in .txt format and were anonymized by the research team at Eurac. At the moment of writing, the corpus is not yet publicly available, and it was obtained directly from Eurac Research upon the signing of an agreement about its use for research purposes.

3.2.2 Data and analyses

For this dissertation, the data from the corpus, together with the word lists used to analyze it, were processed using AntWordProfiler (Anthony, 2023a) and AntConc (Anthony, 2023b). AntWordProfiler is a free software for profiling vocabulary and estimating the complexity of texts, while AntConc is a freeware corpus analysis toolkit used for concordancing and other types of text analysis. The AIWL and Nvdb lists were obtained from OSF.io²³ and GitHub²⁴ respectively; Nvdb was already divided in three files, according to the sections described in the previous chapter (Section 2.3.2). The metadata of the word lists had to be cleared before loading them into AntWordProfiler to start the analyses; Ndvb files contained HTML elements, which were removed. Words in the lists are in alphabetical order, and Nvdb files contains part of speech information, which was ignored for this study.

A major issue concerning the word lists was lemmatization: the words in the lists were present only in their base form. To solve this, I added all the different forms for each lemma, taking them from the morphological dictionary MorphIt!²⁵ (Zanchetta and Baroni, 2005). However, some forms appeared more than once, since in Italian it is common to have the same form for different grammatical categories; for example, the adjective form and the present participle form often match (e.g., *culminante*, ‘culminating’ which is the present participle form of the verb *culminare*, ‘culminate’, but is also used as an adjective). The same goes for verbs, where the same form can be used in different modes and tenses: the first-person plural in the indicative present form is the same as the first-person plural in the imperative form, for instance (e.g. *preghiamo*, ‘pray’). The repetitions inside the single

²³ The resource is available at the following link: https://osf.io/sa8fp/?view_only=2d2e872ae41d47e9b374fba7c96fab60 (Last visited: 19-11-2023).

²⁴ The resource is available at the following link: <https://github.com/memdevice/nvdb> (Last visited: 19-11-2023). Nvdb was originally published as a pdf file, accessible from the following link, which could not be used directly in AntWordProfiler: <https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana> (Last visited: 20-11-2023).

²⁵ A presentation of the corpus is available at the following link: <https://cris.unibo.it/handle/11585/1532> (Last visited: 19-11-2023).

lists were therefore removed. Another issue that I faced with lemmatization was that not all words present in the Nvdb were included in the MorphIt! dictionary; for this reason, since I decided to give the priority to the criterion of frequency, only their base form was considered, also because the inflected forms are not frequent or match the base form; examples of this case, which occurred for nouns and adjectives, include terms like *babydoll*, *aceto* ('vinegar'), *acca* ('h') and *attaccapanni* ('coat rack'). At the beginning, I considered lemmatizing the corpus as a solution to the lemmatization problems, but this was not accurate enough, so I decided to go for a less automatic, but more precise approach.

Going back to repetitions, they were present also across the four lists: in the lists by De Mauro specifically, some of the words appeared in more than one list file, too, i.e., some words belonging to one list were also present in another. This is often due to homonymy, or to some word forms acquiring independent use too (e.g., *accusa* as noun, 'accusation', vs. verb form, 'accuses'); the criteria used for compiling the three lists making up Nvdb, however, sometimes appear mixed: the word *abbandonata* ('abandoned'), for example, appears separately from its corresponding verb *abbandonare* ('to abandon'), in an entry marked both as an adjective and as past participle. The operation of finding duplicates was possible thanks to an automatic function available directly on AntWordProfiler, called "Validate Level Lists"; with its help, it was possible to obtain the words that were present in multiple lists. Given that the tool automatically ignores the words repeated across the lists and it profiles just the form appearing first, the three lists composing the Nvdb were uploaded in order of frequency: first the fundamental vocabulary ("lessico fondamentale", FO), then the high-use vocabulary ("lessico di alto uso", AU) and lastly the high-availability vocabulary ("lessico di alta disponibilità", AD). The decision to follow the criterion of frequency introduces a bias, but my estimate was that this bias would be less pronounced if the less frequent option was omitted.

There were also some repetitions across the Nvdb and the AIWL, due primarily to the fact that AIWL was based on a previous version of Vdb. Spina (2010) designed her list on the model of the English AWL, where terms belonging to general English were excluded; however, the general Italian list has since been updated, leading to some overlaps such as *testo* ('text'), *mercato* ('market') and *società* ('society'). Moreover, these examples come from the FO list, so this means that they are also very frequent in general Italian vocabulary; in the analyses, such words were kept as belonging to the general list.

Concerning academic vocabulary, to be consistent with the decision to focus on frequency, I decided to upload the AIWL as the last reference list for the analysis, also because of the list's nature: it is focused on a specific, more specialized part of vocabulary. In addition, unlike Nvdb, this list contains not only terms, but also collocations: AntWordProfiler, however, does not recognize collocations,

therefore it was not possible to directly identify their presence in the corpus, and N-Grams were studied instead (see below).

Once the .txt files containing the four lists were adjusted, I was finally able to perform the first analyses in AntWordProfiler. After uploading the corpus and the four lists as reference files, I could obtain the percentages related to the coverage of words coming from the four lists that appeared in the corpus. The statistics provided by the software included data for tokens, types, and headwords; I decided to focus on tokens and types, since I believe them to be more interesting also in light of the type/token ratio, which concerns the variety of vocabulary; as explained by Dax (2005), in linguistics the number of tokens corresponds to the total number of words in a text, while the number of types indicates the number of different words. As for the duplicates found by the tool, AntWordProfiler gave as a result 3,541 detected cases; sometimes the repetitions concerned only two lists, rarely three and never all of them; I decided to focus on the repetitions across three lists, since these cases often included AIWL.

With AntConc, I was able to extract the corpus' Wordlist and a list of N-Grams, both based on frequency. I also decided to use AntConc to retrieve some of the multiword expressions of ITACA since it was not possible to obtain data related to the collocations present in AIWL on AntWordProfiler, and the N-Gram option is the only type of complex expression in AntConc that does not need a specific word as input, as opposed to Cluster and Collocate options²⁶.

3.3 Results

In this section, I will first present the descriptive measures and graphs I created for the analyses. Table 4.1 shows the percentages of coverage of the lists in the corpus. The columns to the left show the lists' names and IDs. The ID number 5 referred to the list of terms marked as "ignored" in AntWordProfiler; since no words were included in this list, it was removed from the table. The last list, the one referring to the words that are not present in Nvdb or in AIWL, was automatically created by the software. The data for tokens and types include their number, the corresponding percentages and the cumulative percentages. Since this data directly concerns the research question, i.e., the coverage of the four lists in the corpus, it is also represented graphically, using two bar charts: Figure 4.1 refers to the data on tokens, while Figure 4.2 refers to the data on types.

²⁶ In AntConc, the Cluster function allows to get a list of adjacent word groups if a word is searched as input, while the Collocate option gives a result a list of those words appearing frequently before or after the term chosen; the value of the distance can be set by the user, as well as the direction of the search (left or right to the searched term).

file_name*	list_id	token_count	token_count_%	token_cum_%	type_count	type_count_%	type_cum_%
DeMauro_FO	1	292129	75.34	75.34	6035	42.43	42.43
DeMauro_AU	2	28786	7.42	82.76	2984	20.98	63.42
DeMauro_AD	3	3521	0.84	83.6	462	3.25	66.663
AIWL	4	16440	4.24	87.84	664	4.67	71.33
not_in_lists	6	47166	12.16	100	4077	28.67	100
TOTAL		387772	100		14222	100	

*FO - *Fondamentale* (Fundamental); AU - *Alto uso* (High-use); AD - *Alta disponibilità* (High availability); AIWL - Academic Italian Word List

Table 4.1: Results of the analysis by AntWordProfiler according to tokens and types.

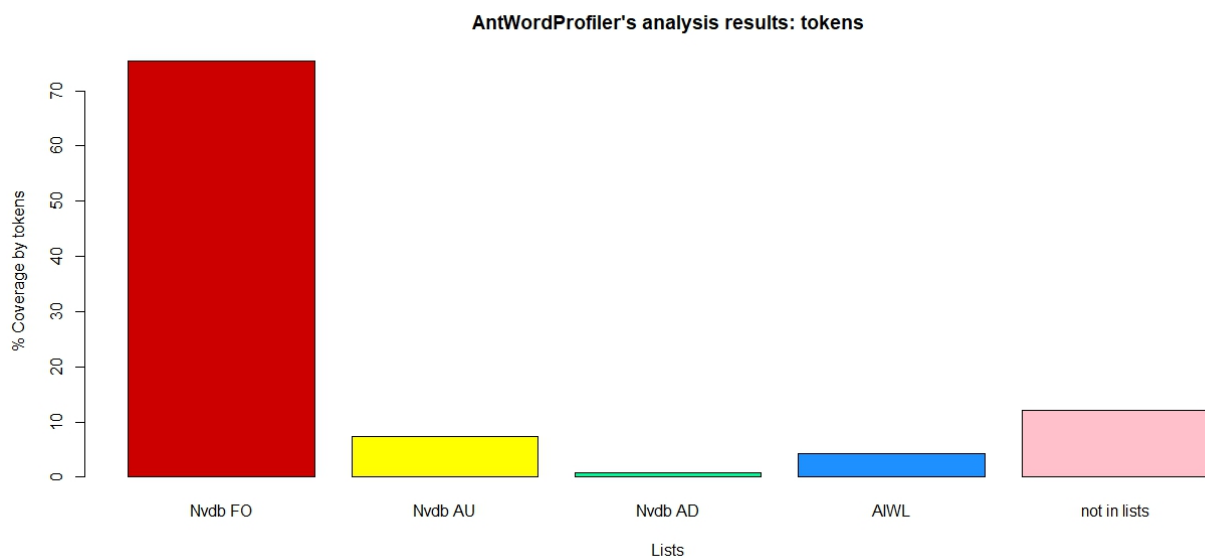


Figure 4.1: AntWordProfiler's analysis results in % (tokens).

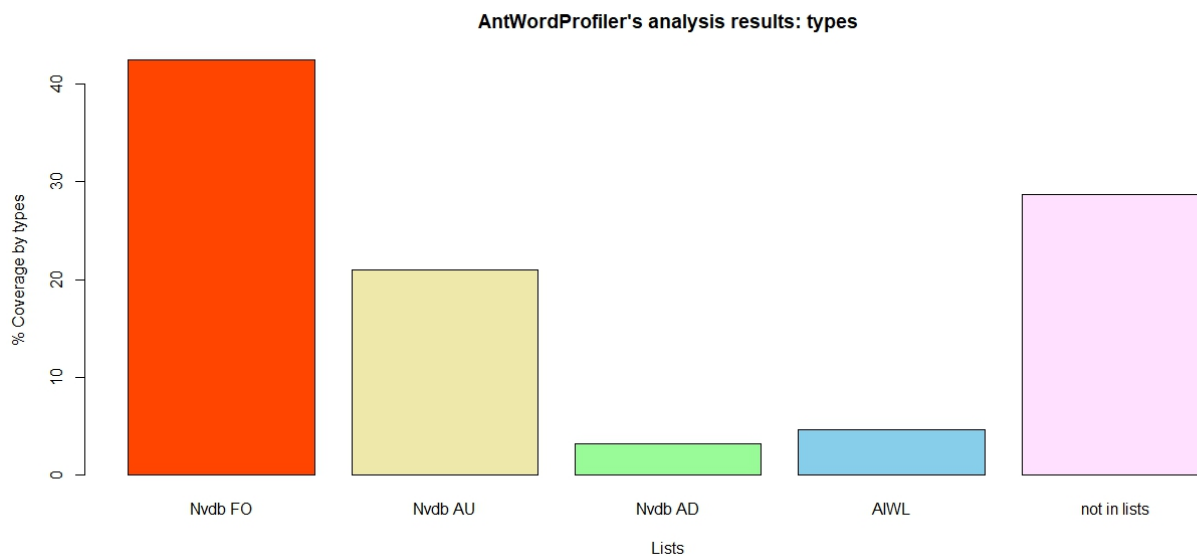


Figure 4.2: AntWordProfiler's analysis results in % (types).

As mentioned earlier, the main focus of this study was trying to see if the Italian word lists chosen for this study – the ones making up De Mauro's Nvdb and Spina's AIWL – are relevant when looking at texts written by young Italian L1 students in academic contexts. When looking at the results measured in tokens, the analysis conducted with AntWordProfiler revealed that over 83% of the terms present in the ITACA corpus come from Nvdb's lists, and around 4% from AIWL. As for the results related to Nvdb specifically, the highest percentage concerns the FO list, which comes as no surprise, since the words contained in this list are the most frequent ones in the Italian language. The remaining two (sub)lists did not show unexpected results, as they follow the criterion of frequency (AU: 7,42%; AD: 0,84).

When examining the results measured by types, there are not many differences from the reference lists: FO's percentage remains the highest, while AIWL's value is almost the same as the level of coverage measured in tokens. An interesting result has to do with the words not in lists, which represent 12% when measured in tokens, more than AU, AD and AIWL taken individually, and more than 28% when looking at type, again the second highest result after FO; this finding may be related to the words concerning the COVID pandemic, which of course were not used before, and they are also strictly related to the topic of the texts, so repetitions of these terms are not a surprise.

The information about the total number of types also allows for the calculation of the type/token ratio, as anticipated above: by dividing the number of types by the number of tokens, we obtain a ratio indicative of the level of lexical variety of a text: the higher the value of TTR, the higher the lexical complexity of the text. The total number of types in ITACA is 14,222, while the tokens are 387,772;

TTR thus equals 3.66%. This percentage is quite low, meaning that the lexical variety in ITACA is not very high: this can be explained by the fact that we are looking at the corpus in its entirety (longer text are by definition more repetitive), and the texts all deal with the same topic, so it is natural that the students repeated the same words a lot, also because there are no synonyms available for some of the terms about the COVID pandemic and its impact on school. Another Italian corpus related to academic language, i.e., acWaC IT (part of the acWaC project; Ferraresi and Bernardini, 2015), which is composed of web pages in Italian crawled from websites of Italian universities and has 354,672 unique words and 15,905,016 words in total, has a TTR that equals 2,22%; by comparing these two results, however, it is important to keep in mind the difference in size.

The second part of the analysis concerns the words appearing in multiple lists: there are 70 examples of terms repeated in three lists out of four. A lot of these cases refer to a single term and include its forms varying according to gender and/or number, such as *avanzata* ('advanced'), which occurs also as *avanzati*, *avanzate* and *avanzati*, and *caratteristica* ('characteristic'), which occurs also in its plural form *caratteristiche*. Almost all of these repetitions concern two lists from the Nvdb (in most cases FO and AU, specifically) and AIWL. The overlapping found in Italian word lists between general and academic vocabulary, whose example include *formula* ('formula'), *percorso* ('path') and *specifico* ('specific'), could be a topic to further develop when it comes to creating or adjusting word lists and also to the definition of academic vocabulary, which we discussed in the previous chapter. Moreover, these forms can belong to two different lemmas: they can refer to the noun or adjective, but also to the related verb, i.e., *formulare* ('to formulate'), *percorrere* ('to travel') and *specificare* ('to specify'). As for *formula* specifically, it is not likely that there is a relationship between the noun and the verb, while for *avanzata* it could be that the forms are related (past participle and adjective). As mentioned earlier, it was not possible to consider the information on the part of speech of the words included in the lists, as AntWordProfiler does not allow to process such information. All cases of repetition are shown in Table A.1, in the Appendix.

Table 4.2 shows the top part (20) of the 100 most frequent words in the corpus; this data was extracted with AntConc's Wordlist function; the full table is included in the Appendix as Table A.2.

Type	Rank	Freq	Range
di	1	12603	634
e	2	11016	633
a	3	9633	634
la	4	9573	634
che	5	9273	634
in	6	7909	633
per	7	6464	632

non	8	5281	630
il	9	4986	629
un	10	4642	627
le	11	4630	632
è	12	4593	615
si	13	3616	617
con	14	3189	621
i	15	3113	617
una	16	3034	604
studenti	17	2954	574
più	18	2938	603
l	19	2761	606
didattica	20	2614	577

Table 4.2: First 20 most frequent words of ITACA.

As expected, the words occurring the most are function words. The Merriam-Webster online dictionary defines a function word as “a word (such as a preposition, auxiliary verb, or conjunction) that expresses primarily a grammatical relationship”; examples from this list include indeed prepositions and conjunctions such as *di* (‘of’), *e* (‘and’), *a* (‘to’), *per* (‘for’). From the column with the data on range (referring to the range of tests containing a given word), we can see that these words appear in almost all texts. The first content word we encounter is *studenti* (‘students’), at the 17th position; this is not surprising, since texts were written by students and the focus of the writings was on school. One of the most peculiar frequent terms which is not to be found neither in Nvdb nor in AIWL, is *dad*, an acronym for *didattica a distanza* (‘distance learning’), placed 26th (see Table A.2); then we find *distanza* (‘distance’), a related term, and *casa* (‘house’), which refers to the fact that students followed courses from their homes. Auxiliary verbs also appear in the list, including *è* (‘is’), *sono* (‘are’), *ha* (‘has’) and *hanno* (‘have’). A position that seems surprising instead was the one for *pandemia* (‘pandemic’), appearing only at the 99th place.

Table 4.3 shows the first 20 most frequent N-Grams (full version with 100 N-Grams is present in the Appendix, as Table A.3).

Type	Rank	Freq	Range
didattica a distanza	1	1296	441
la didattica a	2	681	325
didattica digitale integrata	3	652	376
la didattica digitale	4	451	286
seguire le lezioni	5	390	277
la possibilità di	6	244	181
punto di vista	7	221	168

di seguire le	8	203	171
delle scuole superiori	9	196	176
la dad ha	10	193	164
per quanto riguarda	11	187	116
la maggior parte	12	185	146
problemi di socializzazione	13	184	170
triennio delle scuole	14	183	165
che la didattica	15	173	129
il fatto che	16	169	124
a distanza e	17	167	136
le lezioni in	18	165	116
che la dad	19	162	127
didattica in presenza	20	160	105

Table 4.3: First 20 most frequent N-Grams of ITACA.

As we can see, some of the N-Grams do not carry a complete lexical or grammatical meaning by themselves (they are incomplete phrases and sometimes they end with function words), but others do, for instance those related to the topic of the texts, i.e., online learning (*didattica a distanza*, ‘distance learning’, which has a very high range, and *didattica digitale integrata*, ‘integrated digital learning’). Apart from these constructions that are strictly related to the features of the specific writing task, there are some interesting expressions that could be related to academic vocabulary and that are not present in AIWL, such as *punto di vista* (‘point of view’). I chose to look for N-Grams composed of three elements since this allows to obtain many lexical combinations, such as the ones related to the term *didattica* (‘didactics’), whose high frequency is visible in Table 4.2; moreover, many collocations in AIWL are made up of three elements. As opposed to the most frequent N-Grams composed of three elements in ITACA, collocations in AIWL refer mainly to university settings (*corso di laurea*, ‘degree course’), are specific of academic writing (*breve excursus*, ‘brief excursus’) or illustrate its goals (*elaborare una teoria*, ‘to elaborate a theory’); moreover, many of AIWL collocations are built on auxiliary verbs: there are 23 of them with *avere* (‘to have’) as a basis, and 6 of them with *essere* (‘to be’); finally, the N-Grams reflect the topic of the writing task, whereas AIWL is obviously more neutral in this sense.

The last topic to be addressed here is represented by the words that do not belong to any of the four reference lists. As mentioned earlier, AntWordProfiler does not provide a general overview of those words and we only know their coverage in the corpus, as seen in Table 4.1; however, we can see some of those words by looking at the texts individually. The finding that many of the words that are not present in any of the lists refer to the pandemic, the decisions taken in order to limit its negative

effects, and its consequences on people, did not come as a surprise: in many texts, there are words like *pandemia* ('pandemic'), *quarantena* ('quarantine'), *mascherina* ('mask'), *socializzazione* ('socialization'). There are also some English terms, such as *tablet*, *Netflix* and *multitasking*, and also expressions such as *smart working* (for 'agile working') and *project work*. A rise in the use of English terms in Italian has been evident for many years now, as pointed out, for example, by an article in the Italian newspaper *Il Sole 24 Ore* which mentioned it already in 2014²⁷, in particular in the field of technology; the term *tablet* is also in an example. Words referring more specifically to the academic context are also present; examples include *triennio* ('three-year period', referring to the first three years of Italian high school), *neuroscienze* ('neurosciences'), and *questionario* ('survey'). Other interesting aspects worth mentioning include the frequent use of adverbs, such as *drasticamente* ('drastically') and *permanentemente* ('permanently') – a possible explanation for this could be the students' attempt to describe in a more precise way their feelings and the conditions they had to live in during the pandemic – and the use of terms with a negative connotation, for example *svantaggio* ('disadvantage'), *scomodità* ('inconvenience') and *stravolto* ('upset'), which reflect the students' feelings about the drastic changes they had to face at school and in their life because of the pandemic. There are also terms relating to the everyday life conducted by young pupils, such as *routine*, *coetanei* ('peers') and *spensieratezza* ('lightheartedness'), which can be seen as in contrast to the context of the pandemic.

3.4 Discussion

The results suggest that word lists are useful also for profiling L1 learner vocabulary, as we have seen that they cover a high percentage of the lemmas chosen by young students. The representativeness of Italian word lists is still high even if the topic on which the ITACA texts were based reflects an event that significantly marked the last few years also through vocabulary choices; this also proves the versatility of such a resource. One reason for this is the presence of grammatical words, which were also mentioned earlier and tend not to change much. As for the lexical items instead, apart from the terms related to the pandemic and the words in English, there seems to be more words in this corpus related to emotions and feelings. Still, Nvdb and AIWL are both relevant, even if not directly comparable given their different features and goals; AIWL might be useful also for analyzing writing in university contexts, as vocabulary becomes more specialized with time.

²⁷ Article available at the following link: <https://st.ilsole24ore.com/art/notizie/2014-05-13/call-posto-telefonata-termini-inglesi-spopolano-azienda-ecco-top-ten-piu-utilizzati-104202.shtml?uuiid=AB2MbqHB> (Last visited: 14-11-2023).

As we have said in the previous chapter, there are not many analyses based on vocabulary profiling in Italian; a possible explanation could lie in the morphological complexity of this language, as opposed to English. The morphological complexity of Italian reminds us also of the notion of headword, which is criticized in English despite its use in some word lists, as we explained before, but would possibly be even less useful in Italian. The variety concerning gender, number, and also verb tenses makes it not useful to perform such analyses. For this reason, it is important to also remember the language's nature and characteristics when looking at possible gaps in research.

This study is the demonstration of the possibility of using word lists also when it comes to analyze L1 students writing, and such research could also bring implications for teaching: we can hypothesize that a use of word lists could lead to more vocabulary variety in the writings by L1 students, for example, and to a rise in the pupils' general awareness of their mother tongue, which is important and should not be considered less important than the study and acquisition of new languages: as stated by Kadarisman (2009), language is indeed a mirror of society, and if it is important to know the society we live in, with its cultures, its traditions and its characteristics, then it is vital to not stop the learning and development of native languages. For this reason, it is important to acknowledge all the opportunities offered by tools such as word lists, which might also be considered as a sort of mirror of language use. Depending on the school subject, for example, teachers could take inspiration from the academic word lists and create a specific subset according to their students' needs and to their knowledge level; they could also use word lists to create contents in the pupils' L1, or they could engage children in practical activities connected to all communication skills and not just to writing: word lists could then be seen as a versatile tool and not "just" as a simple and monolithic list of terms to memorize.

Conclusion

As we have seen from the results of the analyses that were conducted, word lists still represent a useful source to understand and analyze vocabulary. Despite their limitations, which we reviewed in Chapter 2, we have seen that in this empirical study the lemmas contained in the word lists are still used today by young people at school. Given the fact that the corpus chosen for this research is composed of texts written by Italian L1 pupils, we could conclude that word lists can be also useful to analyze L1 vocabulary, as the word lists used for this study covered important percentages of the corpus texts.

As for future studies, word lists could then be used to see possible changes in language, too, for example when it comes to measuring the impact of the COVID pandemic: the N-Grams which were presented in the last chapter could be useful, for example, for the analysis of the pandemic's consequences on students.

As for the ITACA corpus itself, it might be useful to analyze its collocations, since it was not possible with the tools that were used in this research: more work on the tools should also be done, in order to better grasp also the cases where collocations are divided by other lemmas in between; it would have been interesting to see the presence of the collocations from AIWL, for instance, or to find new ways to study the multi-word expressions which are proper of this corpus. Another suggestion concerns the words that were not present in the reference lists, which were available only in the profiling of the single texts: a general view of those words could be interesting to analyze. It could also be interesting, for a future study, to compare the results of this thesis with an analysis based on corpus lemmatization. Moreover, an Italian list of collocation in general language, for example, could surely represent a great innovation in vocabulary studies, so that it could be used together with the collocations in AIWL. It would also be important to continue researching and analyzing the way young people write today, also with longitudinal studies, to see if changes occur over time. In general, student corpora like ITACA provide helpful information on many language features, as we have discussed, but unfortunately they are not numerous. It could also be interesting to analyze the level of vocabulary knowledge of L1 students with vocabulary tests, in order to better identify possible problems and difficulties faced by them, so that teachers can use more adequate materials and teaching methods; it is essential, however, to keep in mind the differences in the education in high school compared to university, and the consequently differences in the specialization of the vocabulary.

Vocabulary represents an important part of language: choosing the right words is often the key when communicating, and it is not easy to do so, since words have a lot of different nuances. For this reason,

we need to further develop the study field of vocabulary, also to understand better where society is headed, given the fact that, as stated earlier, language is a mirror of society.

References

- Ackermann, K., Chen, Y.-H. (2013). “Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach”. *Journal of English for Academic Purposes*, 12(4): 235-247.
- Ackermann, K., De Jong, J. H. A. L., Kilgarriff, A., Tugwell, D. (2011) “The Pearson International Corpus of Academic English (PICA-E)”, *Proceedings of Corpus Linguistics*. <https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2011/Paper-47.pdf> [Last visited: 20-11-2023].
- Al Zahrani, S. M., Chaudhary, A. (2022). “Vocabulary Learning Strategies in ESP Context: Knowledge and Implication”. *Arab World English Journal*, 13(1). <https://dx.doi.org/10.24093/awej/vol13no1.25> [Last visited: 20-11-2023].
- Alharbi, A. M. (2019). “Investigating sequential vocabulary learning strategies as a means of improving L2 vocabulary acquisition”. University of Memphis, Electronic Theses and Dissertations. 2408. <https://digitalcommons.memphis.edu/etd/2408> [Last visited: 20-11-2023].
- Anderson, R.C., Freebody, P. (1982). “Reading Comprehension and the Assessment and Acquisition of Word Knowledge”. Technical Report 249, Illinois University.
- Anthony, L. (2023a). AntWordProfiler (Version 2.1.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Anthony, L. (2023b). AntConc (Version 4.2.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Barnes, M. (2012). *An Investigation of English Learners' Acquisition of Academic Language to Obtain Reading Comprehension Skills*. University of Liberty. <https://digitalcommons.liberty.edu/doctoral/610/> [Last visited: 15-11-2023].
- Bayley, R., Ceil, L. (Eds, 2007). *Sociolinguistic Variation: Theories, Methods, And Applications*. Cambridge: Cambridge University Press.
- Bennett, G.R. (2010). *Using corpora in the language learning classroom: corpus linguistics for teachers*. Ann Arbor: University of Michigan Press.
- Berman, R., Cheng, L. (2010). “English academic language skills: Perceived difficulties by undergraduate and graduate students, and their academic achievement”. *Canadian Journal of Applied Linguistics*, 4(1): 25–40. Retrieved from <https://journals.lib.unb.ca/index.php/CJAL/article/view/19830> [Last visited: 15-11-2023].
- Biber, D. (2006). *University Language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins Publishing.
- Bienati, A., Vettori, C., Zanasi, L. (2022). “In viaggio verso ITACA: la coerenza testuale come meta della scrittura scolastica. Proposta di una griglia di valutazione”. *Italiano a scuola*, 4: 55-70. <https://italianoascuola.unibo.it/article/view/14830/15638> [Last visited: 17-11-2023].

- Borawski, G. H. (2019). "Optimizing L2 Vocabulary Acquisition: Applied Linguistic Research". Purdue Linguistics, Literature, and Second Language Studies Conference, 1. <https://docs.lib.purdue.edu/plcc/2019/papers/1> [Last visited: 20-11-2023].
- Bortolini, U., Tagliavini, C. e Zampolli, A. (Eds, 1972). *Lessico di frequenza della lingua italiana contemporanea (LIF)*. Milan: Garzanti.
- Boscolo P., Zuin, E. (Eds, 2015). *Come scrivono gli adolescenti. Un'indagine sulla scrittura scolastica e sulla didattica della scrittura*. Bologna: Il Mulino. 17-21.
- Casañ Pitarch, R. (2016). *Possibilities and limitations in a corpus-based model of genre analysis and future perspectives*.
- Chen, Q., Ge, G.-C. (2007). "A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs)", *English for Specific Purposes*, 26(4): 502-514.
- Christie, F., Derewianka, B. (2008). *School discourse: Learning to write across the years of schooling*. London: Continuum.
- Ciliberti A., Anderson L. (Eds, 1999) *Le forme della comunicazione accademica: ricerche linguistiche sulla didattica universitaria in ambito umanistico*. Milano: Franco Angeli.
- Corino, E., Marellò, C. (2009a). "Didattica con i corpora di Italiano per stranieri". *Italiano LinguaDue*, 1: 279-285.
- Corino, E., Marellò, C. (Eds, 2009b). *VALICO. Studi di linguistica e didattica*. Perugia: Guerra.
- Coxhead, A. (1998). *An academic word list*. E.L.I. Occasional publication, 18. New Zealand: Victoria University of Wellington.
- Coxhead, A. (2011). "The Academic Word List 10 Years on: Research and Teaching Implications". *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect*, 45(2): 355-362.
- Coxhead, A., Nation, P., Sim, D. (2014). "Creating and trialling six versions of the Vocabulary Size Test". *TESOLANZ Journal*, 22: 13-27.
- Cronbach, L. J. (1942). "An analysis of techniques for diagnostic vocabulary testing". *The Journal of Educational Research*, 36(3): 206-217.
- Cummins, J. (1979). "Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters". *Working Papers on Bilingualism*. 19: 121-129.
- Cummins, J. (2008). *BICS and CALP, Encyclopedia of Language and Education*. https://link.springer.com/referenceworkentry/10.1007/978-0-387-30424-3_36 [Last visited: 07-11-2023].
- D'Aguanno, D. (2019). "Il lessico accademico per l'insegnamento della scrittura nelle scuole superiori". In M. Palermo and E. Salvatore (Eds), *Scrivere nella scuola oggi: Obiettivi, metodi,*

esperienze. *Atti del secondo Convegno nazionale ASLI Scuola Siena, Università per Stranieri, 12-14 ottobre 2017*. Florence: Franco Cesati Editore.

Dal Negro, A. (2016). “La competenza passiva dell’Academic Italian Word List”. XIV CercleS International Conference 2016 - Enhancing Learners’ Creative and Critical Thinking: The Role of University Language Centres.

Davies, M. (2008-). *The Corpus of Contemporary American English: 425 million words, 1990-present*. <http://corpus.byu.edu/coca/> [Last visited: 20-11-2023].

Dax, T. (2005). *Type-token Ratios in One Teacher’s Classroom Talk: An Investigation of Lexical Complexity*.

De Mauro, T. (1980). *Vocabolario di base della lingua italiana*. In T. De Mauro, Guida all’uso delle parole. Rome: Editori Riuniti.

De Mauro, T. (2016). *Nuovo vocabolario di base della lingua italiana, Internazionale*. <https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana> [Last visited: 20-11-2023].

Desideri P. (2011), “Pragmatica, argomentazione e glottodidattica del discorso accademico”. In Desideri P. and Tessuto G. (2011), 43-72.

Desideri P., Tessuto G. (2011). *Il discorso accademico: lingue e pratiche disciplinari*, Urbino: Quattroventi.

Duffelmeyer, F.A. (1980). “The influence of experience-based vocabulary instruction on learning word meanings”. *Journal of Reading*, 24(1): 35-40.

Dunn, L. M. (1959). *Peabody Picture Vocabulary Test*. Minneapolis: American Guidance Service.

Durrant, P. (2009). “Investigating the viability of a collocation list for students of English for academic purposes”. *English for Specific Purposes*, 28(3): 157-169.

Durrant, P. (2016). “To what extent is the Academic Vocabulary List relevant to university student writing?”. *English for Specific Purposes*, 43: 49-61.

Elley, W. B. (1989). “Vocabulary acquisition from listening to stories”. *Reading Research Quarterly*, 24(2): 174–187. <https://doi.org/10.2307/747863> [Last visited: 20-11-2023].

Fallani G., Troncarelli D. (2018), “Strategie e risorse per lo sviluppo delle abilità di scrittura accademica”. In E. Jafrancesco (Ed), *Le competenze trasversali dello studente universitario di italiano L2*. Siena: Becarelli. 107-134.

Faucett, L., Palmer, H., Thorndike, E. L., West, M. (1936). *Interim report on vocabulary selection*. London: P.S. King and Son, Ltd.

Feilke, H. (2012). „Bildungssprachliche Kompetenzen fördern und entwickeln” [“Fostering and developing academic language competencies“]. *Praxis Deutsch*, 233: 1–13.

- Ferguson, G., Perez-Llantada, C., Plo, R. (2011). "English as an international language of science publications: a study of attitudes". *World Englishes*, 30: 41-59.
- Ferraresi, A., Bernardini, S. (2013). "The academic Web-as-Corpus". In Evert, S., Stemle, E., Rayson, P. (Eds). *Proceedings of the 8th Web as Corpus workshop (WAC-8)*, Lancaster, UK.
- Ferreri, S. (2005), *L'alfabetizzazione lessicale. Studi di linguistica educativa*. Rome: Aracne.
- Fillmore, C. (1992). "Corpus-Linguistics vs. Computer-Aided Armchair Linguistics". In J. Svartvik (Ed.), *Directions in Corpus Linguistics*: 35-60. Berlin: De Gruyter.
- Fiorentino, G. (2015). "Aspetti problematici del discorso accademico: un'analisi dei riassunti delle tesi di laurea". *Cuadernos de Filología Italiana*, 22: 263-284.
- Flynt, E., Brozo, W.G. (2008). "Developing Academic Language: Got Words?". *The Reading Teacher*, 61(6): 500-502.
- Forti, L., Milani, A., Piersanti, L., Santarelli, F., Santucci, V., Spina, S. (2019). "Measuring Text Complexity for Italian as a Second Language Learning Purposes". In H. Yannakoudakis, E. Kochmar, C. Leacock, N. Madnani, I. Pilán, T. Zesch (Eds), *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Application*. Florence: Association for Computational Linguistics. 360–368.
- Francis, W.N., Kučera, H. (1964, 1971, 1979). *A Standard Corpus of Present-Day Edited American English, for use with Digital Computers (Brown)*. Brown University.
- Gadda, G. (1995). "Language Change in the History of English: Implications for Teachers". In D. B. Durkin (Ed.), *Language issues* (262-273). White Plains, NY: Longman.
- Gallina, F. (2019). "Condizioni di possibilità per un piano di alfabetizzazione lessicale in italiano L2 per studenti internazionali". *Italiano LinguaDue*, 1.
- Gardner, D., Davies, M. (2013). "A New Academic Vocabulary List". *Applied Linguistics*, 35(3): 1-24.
- Gilner, L. (2011). "A primer on the General Service List". *Reading in a Foreign Language*, 23(1): 65-83.
- Gogolin, I., Lange, I. (2011). „Bildungssprache und Durchgängige Sprachbildung“ [“Academic language and intergrated language education“]. In S. Fürstenau and Gomolla, M., *Migration und schulischer Wandel: Mehrsprachigkeit*. Wiesbaden: VS Verlag. 107–128.
- Granger, S. (2008). "Learner Corpora". In A. Lüdeling and M. Kytö (Eds), *Corpus Linguistics. An International Handbook*, 1: 259-275). Berlin: De Gruyter.
- Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. London: Edward Arnold.
- Hardy, J.A., Römer, U. (2013). "Revealing disciplinary variation in student writing: a multi-dimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP)". In *Corpora*, 8(2): 183-207.

- Hulme, R.C. (2018). *Incidental learning of new meanings for familiar words*. University College London.
- Hulstijn, J. (2003). "Incidental and intentional learning". In C. Doughty and Michael Long (Eds), *Handbook of second language acquisition*. Maryland: Blackwell Publishing.
- Johansson, V. (2008). "Lexical diversity and lexical density in speech and writing: a developmental perspective". Lund University, Dept. of Linguistics and Phonetics, *Working Papers*, 53: 61-79.
- Kadarisman, A.E. (2009). *The "Language is a mirror" metaphor and its implication on foreign language teaching*. State University of Malang.
- Kalogeros, B., Serrano, R., Tragant, E. (2023). "Children's L2 vocabulary acquisition through extensive reading". *Sintagma*, 35: 71-83.
- Karalık, T., Merç, A. (2019). "Correlates of listening comprehension in L1 and L2: A meta-analysis". *Eurasian Journal of Applied Linguistics*, 5(3): 353-383.
- Kong, S., Hoare, P. (2012). "The Development of Academic Language Proficiency: Challenges for Middle School Immersion in Hong Kong and Xi'an". *International Education*, 41(2). <https://trace.tennessee.edu/internationaleducation/vol41/iss2/6> [Last visited: 07-11-2023].
- Krashen, S., Brown, C.L. (2007). "What is Academic Language Proficiency?". *STETS Language & Communication Review*, 6(1).
- Lange, K., Matthews, J. (2020). "Exploring the relationships between L2 vocabulary knowledge, lexical segmentation, and L2 listening comprehension". *Studies in Second Language Learning and Teaching*, 10(4): 723-749. <https://doi.org/10.14746/ssllt.2020.10.4.4> [Last visited: 20-11-2023].
- Laufer, B., Nation, P. (2012). "Vocabulary". In S. Gass and A. Mackey (Eds), *The Routledge Handbook of Second Language Acquisition*. New York: Routledge. 163-176.
- Lessard-Clouston, M. (2013). "Word Lists for Vocabulary Learning and Teaching". *The CATESOL Journal*, 24(1): 287-304.
- Li, Y., Qian, D.D. (2010). "Profiling the Academic Word List (AWL) in a financial corpus", *System*, 38(3): 402-411.
- Ligoš, M. (2020). "Continuity of traditional and new tendencies in L1 teaching: A scientific essay on the challenges and projections". *L1-Educational Studies in Language and Literature*, 20: 1-18. <https://doi.org/10.17239/L1ESLL-2020.20.01.11> [Last visited: 15-11-2023].
- Liu, L., Li, L. (2016). "Noun Phrase Complexity in EFL Academic Writing: A Corpus-Based Study of Postgraduate Academic Writing". *The Journal of Asia TEFL*, 13(1): 48-65.
- Martínez, I.A., Beck, S.C., Panza, C.B. (2009). "Academic vocabulary in agriculture research articles: A corpus-based study". *English for Specific Purposes*, 28(3): 183-198.
- Martinez, R., Schmitt, N. (2012). "A Phrasal Expressions List". *Applied Linguistics*, 33(3): 299-320.

- Marx, A., Heppt, B., Henshel, S. (2017). "Listening comprehension of academic and everyday language in first language and second language students". *Applied Psycholinguistics*, 38: 571-600
- Mastrantonio, D. (2021). "L'italiano scritto accademico: problemi descrittivi e proposte didattiche". *Italiano LinguaDue*, 1.
- Meara, P. (1996). *The vocabulary knowledge framework*. Vocabulary Acquisition Research Group, Swansea University. <https://www.lognostics.co.uk/vlibrary/meara1996c.pdf> [Last visited: 20-11-2023].
- Meganathan, P.M., Thai, Y.N., Paramasivam, S., Jalaluddin, I. (2019). "Incidental and Intentional Learning of Vocabulary among Young ESL Learners". *3L: The Southeast Asian Journal of English Language Studies*, 25(4): 51-67.
- Meiners-Lovel, L.A. (2020). *How Preschool Teachers Use Book Sharing Strategies to Build Academic Language Skills*. University of Memphis, Electronic Theses and Dissertations. 2675. <https://digitalcommons.memphis.edu/etd/2675> (Last visited: 07-11-2023).
- Michigan Corpus of Upper-level Student Papers*. (2009). Ann Arbor: The Regents of the University of Michigan.
- Millin, T.J. (2016). *Scaffolding academic literacy using reading to learn methodology: An evaluative study*, Stellenbosch University. <https://scholar.sun.ac.za/> [Last visited: 07-11-2023].
- Milton, J., Fitzpatrick, T. (2017). *Dimensions of Vocabulary Knowledge*. London: Bloomsbury Publishing.
- Morek, M., Heller, V. (2012). "Bildungssprache: Kommunikative, epistemische, soziale und interaktive Aspekte ihres Gebrauchs" ["Academic language: Communicative, epistemic, social, and interactive aspects of its utilization"]. *Zeitschrift für angewandte Linguistik*, 2: 67–101.
- Mutema, F. (2022). *The development of reading literacy skills in the early years of primary schooling: A case of four Zimbabwean schools*. University of South Africa. https://uir.unisa.ac.za/bitstream/handle/10500/29629/thesis_mutema_f.pdf?sequence=1&isAllowed=y [Last visited: 07-11-2023].
- Narvaez-Berthelemot, N., Russel, J.M. (2001). "World distribution of social science journals: A view from the periphery". *Scientometrics*, 51(1): 223-239.
- Nation, P. (1983). "Testing and teaching vocabulary". *Guidelines*, 5(1): 12–25.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, P. (2006). "How large a vocabulary is needed for reading and listening?" *Canadian Modern Language Review*, 63(1): 59–82.
- Nation, P., Beglar, D. (2007). "A vocabulary size test". *The Language Teacher*, 31(7): 9-13.

- Nation, P., Waring, R. (1997). "Vocabulary size, text coverage and word lists". In N. Schmitt and M. McCarthy (Eds), *Vocabulary: Description, Acquisition, and Pedagogy*. Cambridge: Cambridge University Press. 6-19.
- Natividad, L.F. (1999). "Helping students recognize cause-effect relations in scientific texts". In *III Congrés Internacional sobre Llengües per a Finalitats Específiques*. Canet de Mar (Barcelona): Publications of University of Barcelona: 140-144.
- Nesi, H., Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge: Cambridge University Press.
- Nesi, H., Gardner, S. (2013). "A classification of genre families in university student writing". *Applied Linguistics*, 34(1): 25-52.
- O’Keeffe, A., Carter, R., McCarthy, M. (2007). *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University.
- Ortega, S.A.N. (2017). *Comprehending and Speaking about Motion in L2 Spanish: A Case of Implicit Learning in Anglophones*. Cham: Springer International Publishing.
- Pagliara, F. (2018). "L’analisi dei bisogni: uno studio sull’italiano come lingua della comunicazione accademica". In D. Cortés Velásquez and E. Nuzzo (Eds), *Il task nell’insegnamento delle lingue. Percorsi tra ricerca e didattica al CLA di Roma Tre*. Roma: Roma Tre-Press: 77-104.
- Pavičić Takač , V. (2008). *Vocabulary learning strategies and foreign language acquisition*. Clevedon: Multilingual Matters.
- Pugliese, R., Della Putta, P. (2020). "Il discorso accademico scritto degli studenti universitari nelle prove di esame: un confronto tra italiano L1 e L2". *Italiano LinguaDue*, 1.
- Richards, J.C. (1976). "The role of vocabulary teaching". *TESOL Quarterly*, 10(1): 77-89.
- Römer, U., Arbor, A. (2009). "English in Academia: Does Nativeness Matter?". *Anglistik: International Journal of English Studies*, 20(2): 89-100.
- Romero, E., Soria, B. (2005). *The notion of grammatical metaphor in Halliday*. University of Granada.
- Scarcella, R. (2003). *Academic English: A conceptual framework*. Technical Report 2003-1, University of California, Linguistic Minority Research institute.
- Schleppegrell, M.J. (2004). *The Language of Schooling: A Functional Linguistics Perspective*. Abingdon: Taylor & Francis.
- Schleppegrell M.J., O’Hallaron, C.L. (2011). "Teaching Academic Language in L2 Secondary Settings". *Annual Review of Applied Linguistics*, 31.
- Schmitt, N. (2014). "Size and Depth of Vocabulary Knowledge: What the Research Shows". *Language Learning*, 64(4): 913-951.

- Schmitt, N., Schmitt, D. (2014). "A reassessment of frequency and vocabulary size in L2 vocabulary teaching". *Language Teaching*, 47(4): 484-50.
- Schmitt, N., Schmitt, D., Clapham, C. (2001). "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test". *Language Testing*, 18(1): 55–88.
- Sekhar Rao. V.C. (2022). "The Academic Language Used in Scientific Research". *Journal for Research Scholars and Professionals of English Language Teaching*, 6(29).
- Smith, S. (2021). *Academic Vocabulary List (AVL)*, EAP Foundation. <https://www.eapfoundation.com/vocab/academic/other/avl/> [Last visited: 20-11-2023].
- Snow, C. E., Uccelli, P. (2009). "The challenge of academic language". In D.R. Olson and N. Torrance, *The Cambridge Handbook of Literacy*. Cambridge: Cambridge University Press. 112-133.
- Spina, S. (2008). *Academic Italian Corpus*. University for Foreigners of Perugia.
- Spina, S. (2010). "AIWL: una lista di frequenza dell'italiano accademico". Statistical Analysis of Textual Data, Proceedings of the 10th Conference JADT.
- Tong, Y., Hasim, Z. Halim, H.A. (2023). "The relationship between L2 vocabulary knowledge and reading proficiency: The moderating effects of vocabulary fluency". *Humanit Soc Sci Commun*, 10(555). <https://doi.org/10.1057/s41599-023-02058-2> [Last visited: 20-11-2023].
- Uessler, S., Runge, A., Redder, A (2013). „Bildungssprache diagnostizieren: Entwicklung eines Instruments zur Erfassung von bildungssprachlichen Fähigkeiten bei Viert- und Funftklässlern" ["Diagnosis of academic language: Development of an instrument for assessing academic language in four- and fifth-graders"]. In A. Redder and S. Weinert (Eds), *Sprachförderung und Sprachdiagnostik: Interdisziplinäre Perspektiven*. Münster: Waxmann. 4–67.
- Vollmer, H. J., Thurmann, E. (2010). „Zur Sprachlichkeit des Fachlernens: Modellierung eines Referenzrahmens für Deutsch als Zweitsprache" ["Language in learning: Formulating a framework of reference for German as a second language"]. In B. Ahrenholz (Ed), *Fachunterricht und Deutsch als Zweitsprache*. Tübingen: Gunter Narr. 107–132.
- Webb, S. (2013). "Depth of vocabulary knowledge". In C. Chapelle (Ed), *Encyclopedia of Applied Linguistics*, Oxford: Wiley-Blackwell. 1656–1663.
- West, M. (1953). *A General Service List of English Words*. London: Longman, Green and Co.
- Wilson, S., Thorne, A., Stephens, M., Ryan, J., Moore, S., Milton, J., Brayley, G. (2016). "English Vocabulary Size in Adults and the Link with Educational Attainment". *LIF – Language in Focus Journal*, 2.
- Zanchetta, E., Baroni, M. (2005). "Morph-it! A free corpus-based morphological resource for the Italian language". *Proceedings of Corpus linguistics Conference Series 2005*: 1-12.
- Zwiers, J. (2008). *Building academic language: Essential practices for content classrooms*. San Francisco: Jossey-Bass.

Zwiers, J. (2014). *Building Academic Language: Meeting Common Core Standards Across Disciplines - Grades 5-12*. Hoboken: John Wiley & Sons.

Appendix

ID	Entry	Headword(s)*	Level Lists
172	avanzata	avanzata [L4], avanzata [L1], avanzata [L2]	1, 2, 4
173	avanzate	avanzate [L1], avanzate [L4], avanzate [L2]	1, 2, 4
174	avanzati	avanzati [L1], avanzati [L2], avanzati [L4]	1, 2, 4
175	avanzato	avanzato [L1], avanzato [L2], avanzato [L4]	1, 2, 4
221	caratteristica	caratteristica [L2], caratteristica [L1], caratteristica [L4]	1, 2, 4
222	caratteristiche	caratteristiche [L2], caratteristiche [L1], caratteristiche [L4]	1, 2, 4
302	colli	colli [L1], colli [L2], colli [L3]	1, 2, 3
333	condizioni	condizioni [L2], condizioni [L1], condizioni [L4]	1, 2, 4
359	contesti	contesti [L1], contesti [L2], contesti [L4]	1, 2, 4
360	contesto	contesto [L1], contesto [L2], contesto [L4]	1, 2, 4
375	corretta	corretta [L2], corretta [L4], corretta [L1]	1, 2, 4
376	corrette	corrette [L4], corrette [L1], corrette [L2]	1, 2, 4
377	corretti	corretti [L4], corretti [L1], corretti [L2]	1, 2, 4
378	corretto	corretto [L1], corretto [L4], corretto [L2]	1, 2, 4
386	crescente	crescente [L4], crescente [L1], crescente [L2]	1, 2, 4
387	crescenti	crescenti [L2], crescenti [L4], crescenti [L1]	1, 2, 4
418	diffusa	diffusa [L4], diffusa [L2], diffusa [L1]	1, 2, 4
419	diffuse	diffuse [L1], diffuse [L4], diffuse [L2]	1, 2, 4
420	diffusi	diffusi [L4], diffusi [L2], diffusi [L1]	1, 2, 4
421	diffuso	diffuso [L1], diffuso [L4], diffuso [L2]	1, 2, 4
537	esistente	esistente [L2], esistente [L4], esistente [L1]	1, 2, 4
538	esistenti	esistenti [L1], esistenti [L4], esistenti [L2]	1, 2, 4
596	formati	formati [L2], formati [L4], formati [L1]	1, 2, 4
597	formato	formato [L1], formato [L4], formato [L2]	1, 2, 4
598	formula	formula [L4], formula [L1], formula [L2]	1, 2, 4
827	individuale	individuale [L2], individuale [L1], individuale [L4]	1, 2, 4
919	limitata	limitata [L1], limitata [L4], limitata [L2]	1, 2, 4
920	limitate	limitate [L1], limitate [L4], limitate [L2]	1, 2, 4
921	limitati	limitati [L2], limitati [L1], limitati [L4]	1, 2, 4
922	limitato	limitato [L1], limitato [L4], limitato [L2]	1, 2, 4
928	mancata	mancata [L4], mancata [L2], mancata [L1]	1, 2, 4
929	mancate	mancate [L4], mancate [L1], mancate [L2]	1, 2, 4
930	mancati	mancati [L1], mancati [L2], mancati [L4]	1, 2, 4
931	mancato	mancato [L2], mancato [L1], mancato [L4]	1, 2, 4
942	medica	medica [L2], medica [L3], medica [L1]	1, 2, 3
949	medico	medico [L1], medico [L2], medico [L3]	1, 2, 3
956	messe	messe [L2], messe [L3], messe [L1]	1, 2, 3
1010	partecipante	partecipante [L2], partecipante [L4], partecipante [L1]	1, 2, 4
1011	partecipanti	partecipanti [L2], partecipanti [L4], partecipanti [L1]	1, 2, 4
1081	percorso	percorso [L4], percorso [L2], percorso [L1]	1, 2, 4
1207	programma	programma [L2], programma [L4], programma [L1]	1, 2, 4
1208	programmi	programmi [L1], programmi [L4], programmi [L2]	1, 2, 4

1239	recuperi	recuperi [L1], recuperi [L4], recuperi [L2]	1, 2, 4
1240	recupero	recupero [L1], recupero [L4], recupero [L2]	1, 2, 4
1275	ridotta	ridotta [L2], ridotta [L4], ridotta [L1]	1, 2, 4
1276	ridotte	ridotte [L2], ridotte [L1], ridotte [L4]	1, 2, 4
1277	ridotti	ridotti [L2], ridotti [L1], ridotti [L4]	1, 2, 4
1278	ridotto	ridotto [L4], ridotto [L2], ridotto [L1]	1, 2, 4
1326	sali	sali [L1], sali [L2], sali [L3]	1, 2, 3
1372	sapiente	sapiente [L3], sapiente [L2], sapiente [L1]	1, 2, 3
1373	sapienti	sapienti [L1], sapienti [L2], sapienti [L3]	1, 2, 3
1479	sorta	sorta [L1], sorta [L2], sorta [L4]	1, 2, 4
1493	specifica	specifica [L2], specifica [L4], specifica [L1]	1, 2, 4
1500	specifico	specifico [L1], specifico [L2], specifico [L4]	1, 2, 4
1524	stabilissimo	stabilissimo [L4], stabilissimo [L2], stabilissimo [L1]	1, 2, 4
1541	supporti	supporti [L1], supporti [L4], supporti [L2]	1, 2, 4
1542	supporto	supporto [L1], supporto [L4], supporto [L2]	1, 2, 4
1544	sveglia	sveglia [L3], sveglia [L1], sveglia [L2]	1, 2, 3
1547	sviluppata	sviluppata [L1], sviluppata [L2], sviluppata [L4]	1, 2, 4
1548	sviluppate	sviluppate [L2], sviluppate [L4], sviluppate [L1]	1, 2, 4
1549	sviluppati	sviluppati [L4], sviluppati [L1], sviluppati [L2]	1, 2, 4
1550	sviluppati	sviluppati [L1], sviluppati [L2], sviluppati [L4]	1, 2, 4
3377	tassi	tassi [L2], tassi [L3], tassi [L4]	2, 3, 4
3378	tasso	tasso [L3], tasso [L2], tasso [L4]	2, 3, 4
1610	utilizzi	utilizzi [L2], utilizzi [L1], utilizzi [L4]	1, 2, 4
1611	utilizzo	utilizzo [L2], utilizzo [L4], utilizzo [L1]	1, 2, 4
1612	vari	vari [L1], vari [L2], vari [L4]	1, 2, 4
1613	varia	varia [L1], varia [L4], varia [L2]	1, 2, 4
1614	vario	vario [L1], vario [L2], vario [L4]	1, 2, 4
1621	verifica	verifica [L1], verifica [L4], verifica [L2]	1, 2, 4

*L1/1 – Nvdb FO; L2/2 – Nvdb AU; L3/3 – Nvdb AD; L4/4 - AIWL

Table A.1: Words appearing in three lists out of the four in total.

Type	Rank	Freq	Range	NormFreq	NormRange
di	1	12603	634	32.501.057	0.997
e	2	11016	633	28.408.446	0.995
a	3	9633	634	24.841.917	0.997
la	4	9573	634	24.687.187	0.997
che	5	9273	634	23.913.537	0.997
in	6	7909	633	20.396.006	0.995
per	7	6464	632	16.669.589	0.994
non	8	5281	630	13.618.828	0.991
il	9	4986	629	12.858.071	0.989
un	10	4642	627	11.970.952	0.986
le	11	4630	632	11.940.006	0.994
è	12	4593	615	11.844.589	0.967
si	13	3616	617	9.325.067	0.970
con	14	3189	621	8.223.905	0.976
i	15	3113	617	8.027.913	0.970
una	16	3034	604	7.824.185	0.950
studenti	17	2954	574	7.617.879	0.903
più	18	2938	603	7.576.617	0.948
l	19	2761	606	7.120.163	0.953
didattica	20	2614	577	6.741.075	0.907
anche	21	2465	592	6.356.828	0.931
sono	22	2394	605	6.173.731	0.951
gli	23	2319	584	5.980.318	0.918
da	24	2308	598	5.951.951	0.940
della	25	2247	589	5.794.642	0.926
dad	26	2234	496	5.761.117	0.780
come	27	1961	582	5.057.095	0.915
scuola	28	1934	547	4.987.467	0.860
distanza	29	1879	517	4.845.631	0.813
ha	30	1817	564	4.685.743	0.887
questo	31	1775	549	4.577.432	0.863
o	32	1703	529	4.391.756	0.832
lezioni	33	1685	527	4.345.337	0.829
ma	34	1660	549	4.280.866	0.863
ad	35	1615	530	4.164.819	0.833
del	36	1597	541	4.118.399	0.851
delle	37	1459	533	3.762.520	0.838
casa	38	1437	519	3.705.786	0.816
ci	38	1437	506	3.705.786	0.796
molto	40	1424	485	3.672.261	0.763
dei	41	1412	529	3.641.315	0.832
presenza	42	1411	512	3.638.736	0.805
al	43	1373	531	3.540.740	0.835
degli	44	1300	477	3.352.485	0.750

alla	45	1298	529	3.347.328	0.832
perché	46	1275	425	3.288.015	0.668
essere	47	1271	510	3.277.699	0.802
hanno	48	1261	507	3.251.911	0.797
se	49	1234	495	3.182.282	0.778
sia	50	1147	438	2.957.924	0.689
problemi	51	1127	460	2.906.347	0.723
tempo	52	1116	494	2.877.980	0.777
nel	53	1070	472	2.759.353	0.742
dell	54	975	443	2.514.364	0.697
tutti	55	969	420	2.498.891	0.660
questa	56	951	433	2.452.472	0.681
può	57	926	419	2.388.001	0.659
digitale	58	913	437	2.354.476	0.687
durante	59	907	403	2.339.003	0.634
molti	60	905	404	2.333.846	0.635
lo	61	889	431	2.292.584	0.678
fare	62	884	390	2.279.690	0.613
modo	63	876	422	2.259.059	0.664
lezione	64	859	373	2.215.219	0.586
ddi	65	853	292	2.199.746	0.459
loro	66	836	382	2.155.906	0.601
mi	67	833	345	2.148.170	0.542
uno	68	829	429	2.137.854	0.675
anni	69	818	423	2.109.487	0.665
apprendimento	70	792	359	2.042.437	0.564
era	71	786	296	2.026.964	0.465
solo	72	778	388	2.006.334	0.610
quanto	73	754	342	1.944.442	0.538
parte	74	753	403	1.941.863	0.634
integrata	75	752	410	1.939.284	0.645
secondo	76	750	379	1.934.126	0.596
verifiche	77	729	398	1.879.971	0.626
scolastico	78	728	388	1.877.392	0.610
quindi	79	714	338	1.841.288	0.531
noi	80	701	301	1.807.763	0.473
fatto	81	697	370	1.797.448	0.582
cui	82	690	359	1.779.396	0.564
stato	83	685	354	1.766.502	0.557
mia	84	681	349	1.756.187	0.549
seguire	85	677	363	1.745.871	0.571
attività	86	673	322	1.735.556	0.506
ed	87	672	349	1.732.977	0.549
nella	88	669	391	1.725.241	0.615
classe	89	652	361	1.681.400	0.568
studio	90	650	345	1.676.243	0.542
tutto	91	646	358	1.665.927	0.563

ho	92	637	320	1.642.718	0.503
anno	93	614	347	1.583.405	0.546
studente	94	610	306	1.573.089	0.481
ore	95	603	319	1.555.037	0.502
ragazzi	95	603	243	1.555.037	0.382
computer	97	598	344	1.542.143	0.541
tra	98	592	350	1.526.670	0.550
pandemia	99	591	343	1.524.091	0.539
dal	100	578	357	1.490.567	0.561

Table A.2: First 100 most frequent words of ITACA.

Type	Rank	Freq	Range	NormFreq	NormRange
didattica a distanza	1	1296	441	3353152	693
la didattica a	2	681	325	1761957	511
didattica digitale integrata	3	652	376	1686925	591
la didattica digitale	4	451	286	1166876	450
seguire le lezioni	5	390	277	1009050	436
la possibilità di	6	244	181	631303	285
punto di vista	7	221	168	571795	264
di seguire le	8	203	171	525224	269
delle scuole superiori	9	196	176	507113	277
la dad ha	10	193	164	499351	258
per quanto riguarda	11	187	116	483827	182
la maggior parte	12	185	146	478652	230
problemi di socializzazione	13	184	170	476065	267
triennio delle scuole	14	183	165	473477	259
che la didattica	15	173	129	447604	203
il fatto che	16	169	124	437255	195
a distanza e	17	167	136	432081	214
le lezioni in	18	165	116	426906	182
che la dad	19	162	127	419144	200
didattica in presenza	20	160	105	413969	165
capacità di seguire	21	150	143	388096	225
gli studenti hanno	22	147	129	380334	203
della didattica a	23	145	102	375160	160
lezioni in presenza	24	143	99	369985	156
in didattica a	25	137	102	354461	160
al triennio delle	26	135	122	349287	192
per gli studenti	27	134	103	346699	162
organizzazione del tempo	28	133	127	344112	200
in presenza e	29	130	111	336350	175
a proprio agio	30	129	116	333763	182
della didattica digitale	30	129	109	333763	171

ridotta capacità di	30	129	126	333763	198
scuola in presenza	30	129	100	333763	157
da input digitali	34	128	125	331176	197
di stanchezza e	35	126	124	326001	195
deficit di attenzione	36	125	120	323414	189
neuroni in modo	36	125	123	323414	193
rispetto a quelle	36	125	123	323414	193
di attenzione e	39	124	118	320826	186
la mia opinione	39	124	102	320826	160
segnali di stanchezza	39	124	121	320826	190
in modo irreversibile	42	122	120	315652	189
ad uno schermo	43	120	99	310477	156
scomodità degli orari	43	120	116	310477	182
che gli studenti	45	119	90	307890	142
le verifiche in	45	119	113	307890	178
di socializzazione e	47	118	117	305302	184
quelle in presenza	47	118	116	305302	182
perdere neuroni in	49	117	115	302715	181
può perdere neuroni	49	117	116	302715	182
a distanza è	51	116	92	300128	145
egregio signor ministro	51	116	114	300128	179
a quelle in	53	114	112	294953	176
andare a scuola	53	114	88	294953	138
con la dad	55	113	96	292366	151
digitali può perdere	55	113	112	292366	176
stanchezza e scarsa	55	113	113	292366	178
di estendere permanentemente	58	112	107	289779	168
e scarsa concentrazione	58	112	112	289779	176
stimolato da input	58	112	110	289779	173
ansia rispetto a	61	111	109	287191	171
input digitali può	62	110	109	284604	171
il multitasking cognitivo	63	109	106	282017	167
con altri impegni	64	108	107	279429	168
l'organizzazione del	64	108	102	279429	160
parte degli studenti	66	107	92	276842	145
con minore ansia	67	106	105	274255	165
dal prossimo anno	68	105	99	271667	156
attenzione e memoria	69	104	102	269080	160
si riesce a	70	103	98	266493	154
a partire dal	71	102	94	263905	148
il cervello se	71	102	102	263905	160
sovrapposizione con altri	71	102	101	263905	159
diminuzione del vocabolario	74	101	99	261318	156
l'utilizzo delle	74	101	98	261318	154
vissute le verifiche	74	101	100	261318	157
tutti gli studenti	77	100	85	258731	134
utilizzo delle tecnologie	77	100	97	258731	153

concentrazione problemi di	79	99	99	256144	156
dedicato allo studio	79	99	95	256144	149
proposta di estendere	79	99	96	256144	151
la dad è	82	98	83	253556	131
scarsa concentrazione problemi	82	98	98	253556	154
il degli studenti	84	97	92	250969	145
verifiche in dad	84	97	93	250969	146
a distanza non	86	96	85	248382	134
agio a casa	86	96	91	248382	143
apprendimento per deficit	86	96	94	248382	148
caduta dell apprendimento	86	96	94	248382	148
e ridotta capacità	86	96	96	248382	151
la scomodità degli	86	96	94	248382	148
famiglie degli studenti	92	94	93	243207	146
tempo dedicato allo	92	94	91	243207	143
dell informazione e	94	92	89	238032	140
minore ansia rispetto	94	92	92	238032	145
per deficit di	94	92	90	238032	142
cervello se ripetutamente	97	91	90	235445	142
dell apprendimento per	97	91	89	235445	140
in grado di	97	91	76	235445	119
lezioni a distanza	97	91	75	235445	118

Table A.3: First 100 most frequent N-Grams of ITACA.