

NOUN PHRASES IN THE INTERLANGUAGE OF CHINESE UNIVERSITY STUDENTS LEARNING ITALIAN AS A SECOND FOREIGN LANGUAGE: A CORPUS STUDY

Shanshan Qu¹, Silvia Scolaro²

1. INTRODUCTION

This paper presents the results of the first longitudinal corpus study conducted over a two-year period with a group of 21 students of Chinese nationality. When the research took place the students were attending the first and second years of the Major course in Italian Language at the Shanghai International Studies University (henceforth referred to as SISU), in the People's Republic of China. The study analyses their oral productions: for each student 23 structured oral productions³ were recorded during the second semester of the first academic year and the first semester of the second academic year along with 2 exam recordings, one final for each year.

All the students participating in the study gave their consent to the use of the recordings anonymously for research purposes.

After collecting the recordings, they were transcribed using the Speechnotes software⁴ and then the transcriptions were manually double checked by the authors using the system developed by Gail Jefferson as described in Sacks, Schegloff and Jefferson (1974). Successively, data were uploaded into a software for speech analysis, Sketchengine.eu⁵, where the noun phrases (NP) were selected, tagged and later analysed by the authors⁶.

Table 1. *Timeline of the present study*

Date	Phase of study	Software used
March 2021- August 2022	Recording collection and transcription	Speechnotes
September 2022	Corpus building and NP tagging	Sketchengine.eu
October-December 2022	Error tagging and analysing	Sketchengine.eu

The following research questions arose during the analysis:

¹ Shanghai International Studies University, R.P.C.

² Università Ca' Foscari, Venice, Italy.

³ Some recordings are missing. See appendix for a complete list including names of students, time and number of tokens of each recording.

⁴ <https://speechnotes.co/#>.

⁵ <https://www.sketchengine.eu/>.

⁶ The software selected the NPs that were later analysed one by one by the authors, who tagged the correct NP. The others were discarded and not counted as "PoS error".

- How does the interlanguage of the noun phrase of the students who were sampled change over the two-year period? Is it possible to find regularities according to second language acquisition theories? If so, are those regularities similar to Chinese learners of Italian in Italy?
- What are the main differences between structured and prepared tasks and unstructured and unprepared tasks? Is it possible to see a difference from one year to the next? In which aspects?

This study can be defined as a bottom-up research: where the study was done at the field level and the analysis of the results found could help refine the theory on language acquisition.

Finally, the empirical evidence delayed pedagogical implications and therefore use (henceforth referred as DPU) could be drawn from the learners' corpora (LC) study's results (Leech, 1998; Gass, Selinker, 2001; Andorno, Rastelli, 2009:7; Granger, 2009; Cacchione, Borreguero Zuloaga, 2017; Gráf, 2017; *inter alia*) to benefit similar-type learners.

2. THEORETICAL UNDERPINNINGS

2.1. *Acquisitional linguistics*

Acquisitional linguistics (AL) is a branch of applied linguistics which studies the development of competence in a second or foreign language. It also aims to define models to explain this process and examines the factors that could have an influence on the procedure of learning categories and structures from an empirical evidence standpoint.

The fundamentals of this branch can be briefly summarized as follows. An important basic concept in this field is that of “interference” (Weinreich, 1974, referring to bilingualism) which can be explained as the divergence from the rules of a certain language caused by the L1, i.e., at a phonological level. Other early contributions to acquisitional linguistics were those of Lado (1957) and his contrastive analysis, along with Chomsky (1959) in his observations about the learning process, in which, instead of considering the learners' output to be full of errors, he posited that the defective output of learners is the result of a progression, a temporary process comprised of changing between new “grammars”, with the learners' productions no longer to be considered full of errors, distant from the LT, but in a change of point of view, the defective output of learners is seen as a result of temporary, in progress, continually changing new ‘grammars’. Corder (1967) follows the same line of thinking: seeing the errors of the learners as a manifestation of different rules that s/he is applying in the process of language learning, i.e., a transitory competence. Selinker (1972) finally labels this phenomenon as “interlanguage” (IL), ushering in AL in the process. The so-called “Morpheme studies” (Dulay, Burt, Krashen, 1985) also played a fundamental role in the development of AL.

Both L1 and L2 participate in the process of Second Language Acquisition (SLA). According to the principles of Universal Grammar (UG) (Chomsky, 1981; Cook, 1990; *inter alia*), structures that are more in compliance with UG, will be learnt more easily, while according to the Linguistic Universals (Greenberg, 1976), less marked structures are those learnt earlier. Additionally, the quantity and quality of input has an influence on the process of SLA that cannot be undervalued.

Research on SLA shows that all learners, independently of L1, go through acquisitional phases in a sequential order, learning some structures before others, starting with language varieties that favour pragmatic and semantic principles, towards varieties where the syntax

of the TL has a more predominant role (Klein, Perdue, 1992; Perdue, 1993; Bernini, 2003; Giacalone Ramat, 2003). In the pre-basic phase, the sentences are short and elementary and are often comment-based NP, without the presence of verbs, making use of intonation to compensate for the lack of linguistic means. In this stage the learners have little communicative autonomy and often rely on the listeners for understanding. In the basic variety, learners are more autonomous on a communicative level, with verbs used more often, even if they are not in their flexed forms, and lexical strategies used to make up what in the TL is expressed with suffixes. Sentences tend to follow the order topic-comment/focus to highlight what is more salient. Sentences are juxtaposed, with little use of subordination. In the post-basic phase, sentences are structured and the syntax starts to be relevant, with verbs tendentially flexed and more complex morphology. This is the phase when suffixes with morphological and syntactical value begin to be more accurate. The phase, although broad, can be divided into intermediate states, advanced varieties and quasi-native ones. It must be noticed that in the progression from one stage to the next during the acquisition phases, there are moments in which structures from the preceding one coexist with those of the next. It can be said that in the IL of the learners, diachronic, but also inter-subjective and interpersonal variations during the process of SLA can be seen, and this variability is what makes the development of the IL possible (Ellis R., 1994b: 119-158). What Ellis R. (1994b) proposes could be detrimental to acquisitional theories but this discussion goes beyond the aim of this paper.

2.2. *Corpus linguistics*

Brezina and McEnery (2021: 11) define corpus linguistics as «an approach to the study of language that uses computers to analyse large amounts of language data, both written and spoken, [called] corpora». Corpus studies have spread recently due to the understanding of both their theoretical and practical benefits (Granger, 2002; 2012) among which at least the following should be mentioned: first of all, the studies are based on data, not on random assumptions; secondly, the data are empirical and thirdly these studies use mass quantities of data, thus making the results more objective and reliable (Brezina, McEnery, 2021). They also allow comparisons to be made between different languages' or between native speakers' (NS) and non-native speakers' (NNS) corpora and the possibility of sharing the data, making them available for researchers all over the world. Scholars (Callies, Götz, 2015; Gilquin, 2015; Nesselhauf, 2004; *inter alia*) have defined language corpora as those produced by learners of a foreign or second language and therefore showing the use of language of the students in that specific moment, representing the learners' IL (Nemser, 1971; Selinker, 1972). In this sense they can be very useful in SLA studies highlighting frequency and distribution of those patterns that the researchers are examining in the sample chosen.

Speaking about “authenticity” which is one of the main characteristics that a corpus should have, in Learners Corpora Research (LCR), there is a debate on the very concept of “authenticity” of the texts, written or spoken, used to collect the various corpora, nonetheless this controversy goes beyond the aim of this study. For the purpose of this research, the materials constituting the corpora will be considered “authentic”, according to Callies and Götz (2015), Granger (2002) and Gilquin and Gries (2009) who consider any texts made by learners in the context of SLA as authentic. Ultimately, Iurato (2022: 714-715) summarizes the definition of learner corpus as «a collection of computerized continuous, spontaneous, contextualized, representative (near-)natural (written or spoken) data produced by foreign or L2 learners, and gathered through those activities

which are ordinarily carried out during the process of teaching and learning of second languages».

2.3. *Spoken Italian corpora*

There are a fair amount of corpora for Italian as a second language. For the purpose of this research, only those relevant to the present study about spoken Italian will be mentioned.

The first corpus-based study is the so-called “Pavia corpus” which was implemented starting in 1986 with the Pavia Project (University of Pavia, Italy) and aimed at observing and analysing the development of spoken interlanguage of speakers with a very distant L1 (Tigrinya, Arabic, Chinese, etc.) in a non-structured context. This corpus is therefore considered not usable for comparison in that it was aimed at observing the IL in spontaneous interactional situations and the sociolinguistic characteristics (age, education, motivation, etc.) of the learners are quite different from those of this research. Even so, looking at it from an acquisitional point of view, some similarities can be found in the IL of Chinese L1 speakers (§ 2.4), i.e., pertaining to the slow acquisition in the use of articles and the prepositions with the articles (Andorno, 2010: 98) as these traits can be seen as interferences from the L1 where articles are absent.

Successively both Perugia and Siena Foreign University (Italy) gathered their corpora from written and oral productions produced during the certification exams held by their university, respectively CELI⁷ and CILS⁸ (Spina *et al.*, 2006; Barni, Gallina, 2008, 2009). It was, however, not possible to use these corpora for comparison due to the inhomogeneity of their contents. Perugia and Siena Foreign universities’ corpora contain exam productions but the tasks given are not comparable to those present in this research.

More recently, in 2019, a corpus of spontaneous spoken Italian in different communicative situations was collected by scholars of the Universities of Bologna and Torino (Italy) and is freely available online (Mauri *et al.*, 2019). Nonetheless, in the present research, the context of the recording did not include spontaneous communicative interactions and therefore a comparison is not completely applicable.

One more corpus that should be acknowledged here is that created at the Sichuan International Studies University of Chongqing in the People’s Republic of China, gathered between September 2018 and January 2019 (Leto, 2023). Despite this corpus’ informants additionally being Chinese university students of Italian, unlike the corpus of the present study, it congregated written, not oral, productions.

2.4. *The noun phrase (NP)*

The noun phrase (NP) is defined by Biber *et al.* (1999: 97) in two ways: firstly the scholars refer to it as «a noun as a head, either alone or accompanied by determiners, and modifiers»; secondly, an NP broadly means «any unit which appears in the positions characteristic of noun-headed structures (including clauses)». In the broad scope, NPs include three major types of constructions: (1) noun-headed phrases; (2) pronoun-headed phrases; (3) nominal clauses, i.e., constituents that are used in the position of a subject, subjective predicative, or direct object (Biber *et al.*, 1999; Hunston, Francis, 2000). In this

⁷ Cf:

<https://www.unistrapg.it/it/certificati-di-conoscenza-della-lingua-italiana/celi-certificati-di-lingua-italiana>.

⁸ <https://cils.unistrasi.it/home.asp>.

study the stricter definition of NP was the one utilized, therefore only referring to noun-headed phrases. In particular NP was chosen as the focus of this research for the peculiar characteristics it holds in Italian compared to the Chinese language, therefore showing it to be acquired slowly. In fact, while in Chinese⁹, a language that belongs to the Sino-Tibetan language family, the noun itself cannot clearly reflect singular or plural information, nor does it have flexions for gender or case, which need to be reflected by the noun phrase where it is located, in Italian the noun has markers of gender and number, it is usually the head of the noun syntagm and it also influences other speech parts connected to it such as articles, adjectives, etc. (Serianni, 1988; Dardano, Trifone, 1995; *inter alia*).

Holmes and Segui (2006: 5) defined gender as follows: «gender describes the syntactic phenomenon according to which some words fall into different classes, such that when they appear in a sentence, other syntactically associated words have to change their form depending upon the specific class». To define the gender of a noun, two main systems can be used: the first being the semantic one connected with the meaning of the noun, while the second is based on formal criteria concerning the morphological and/or the phonological characteristics of the noun (Corbett, 1991: 8). Different languages mix the two systems in various ways. In Italian they tend to overlap. All nouns have an obligatory mark for gender as well as for number and they are expressed in the same suffix which carries both and it fuses together with the root of the noun. This is in agreement with Greenberg's Universal 36: «If a language has the category of gender, it always has the category of number» (Greenberg, 1963: 95). Being that the number category is less marked and more widespread among languages in the world, it should cause fewer difficulties in its acquisition. Nonetheless, this process can be slower for those learners whose L1 nouns do not carry a mark for gender, such is the case of the Chinese. This is confirmed by the analysis done in the Pavia Project corpus (Chini, Ferraris, 2003) where, in the first recordings, the plurals are overused or learners use a lexical strategy (the use of number or a quantifier) to indicate the plural omitting the specific suffix.

Classifying nouns by gender, Chini (1995: 81)¹⁰ divided nouns into seven classes, of which only the first two classes are not ambiguous, expressing gender in a univocal way (-o for masculine singular nouns; -a for feminine singular ones; it should be noted that not all nouns ending with the morpheme -o are masculine and singular, i.e., *mano*: feminine singular). Thus showing that the assignment of gender in Italian is not transparent. This is a relevant point given that the head-noun gender and number is reflected on the other elements in the NP, both internal (adjectives, articles, demonstratives, quantifiers, etc.) and external (past participles, predicative adjectives, etc.) (Renzi *et al.*, 1988; Salvi, Vanelli, 2004). In fact the acquisition of gender in Italian extends through different levels: morpho/phonological, syntactic and semantic (Chini, Ferraris, 2003; Chini, 2005). As far as the morpho-phonological level goes, the learner has to understand that some morphemes convey certain meanings and when associated with an adjective, the latter has to follow the same gender and number categories even when the morpheme carrying that meaning has another form belonging to another class (i.e., *casa piccola*, *casa grande*). On the syntactic level, the learner has to comprehend that the noun functions as a head of the syntagm/phrase and therefore all the variable parts of speech related to it must change accordingly. On the semantic level students discover that gender does not necessarily imply a conceptual gender mark, but it is merely a formal category. This highlights how

⁹ Here for Chinese, it is intended to be putonghua (普通话), the official language of the Chinese Republic of China.

¹⁰ The same classification can be found in Chini, Ferraris (2003: 39).

complex the acquisition of the correct use of gender in the Italian language is for a foreign or second language learner. This is also due to the scarce salience of these morphemes.

Moreover, the noun morphosyntax in Italian also considers the concept of definitiveness ('definitezza') of the noun, which impacts the choice of the definite or indefinite article, and also differs in gender and number. In some cases the article is the only part of speech that enables the understanding of the number and gender of the noun it is associated with, thus making it a clue to support the acquisition of those categories in the Italian language (i.e., *lo sport, il problema, i bar*, etc.). Nevertheless gender, number and determination have different linguistic levels. The gender has its roots on a lexical level: in fact every noun in Italian must fall into one of two classes, masculine or feminine. The criteria followed for this diversification can be semantic, morphological or morpho-phonological (Marcantonio, Pretto, 1988; Serianni, 1988; *inter alia*). Number has a referential value, being a binary category and referring to the contradistinction between singular and plural, except for the *pluralia tantum*¹¹ and the collective nouns¹². The category of definiteness is not flexive in Italian, nor is it based on the grammatical or lexical levels, instead connected with textuality and pragmatics.

In Chini and Ferraris (2003: 41) the rules and regularities concerning the NP in the Italian language are summarized as follow:

- every noun belongs to a flexive class and gender (masculine or feminine);
- nouns, in the majority of cases, are preceded by an article (definiteness);
- the existence of different paradigms of flexive forms for the nouns;
- the arbitrariness of the assignation of gender;
- the low salience of the suffixes for gender and number;
- the fusive and flexive characteristic of the morphology of the NP in Italian;
- the homonymity of some morphemes;
- the differences about the categories of gender, number and definiteness in the L1 of the learner.

Studies on the acquisition of gender for nouns in Italian as a second/foreign language started at the end of the 80's (Marcantonio, Pretto, 1988; Serianni, 1988; Beretta, 1990; Valentini, 1990; *inter alia*) showing that for pre-basic learners there seems to be an overextension of the most frequent allomorphs. In Valentini's (1990) research, where learners were L1 Chinese immigrants to Italy with low education levels, it is found that the ambiguous morphemes are often avoided and there is an overextension of the suffix *-a*, probably perceived as typical of the Italian language by the learners in the study. Moreover, Valentini (1990: 343) states that it seems that the natural gender is stronger than the grammar gender and her research also shows that only in the last recordings does the agreement in the NP start to appear.

Chini and Ferraris (2003: 47), analysing the IL in the productions of different L1 learners, propose the following criteria sequence for the assignment of gender in the Italian language:

morphological (*-o/-a*) → semantic (gender of the referent) → derivational morphology (suffixes).

Learners favour salient and more frequent means, such as *-a* for the feminine singular; followed by less redundant and transparent ones.

¹¹ Nouns which only have the plural form, i.e., *gli occhiali*.

¹² Nouns which express a plural entity, i.e., *la gente*.

Later, thanks to the Pavia Project (Chini, 2005; Giacalone Ramat, 2003), research on the acquisitional steps of Italian language learning starts and develops further. Chini (1995: 285) proposes the following acquisitional sequence:

Anaforico di 3° sg. → articolo determinativo → articolo indeterminativo →
 aggettivo attributivo → aggettivo predicativo → participio passato
 (Anaphoric for third person singular → determinative article →
 indeterminate article → attributive adjective → predicative adjective →
 past participle)

Therefore in the IL of the learners of Italian as a second or foreign language, the NP gender agreement is often one of the features that takes more time to be acquired. When the article is taken into consideration, the determinative article seems to be acquired before the indeterminate one. Moreover, at the basic level, in some cases learners tend to omit the article so as to avoid the problem of flexing it, other times they use a “nativizing” strategy (Andersen, 1983), making the article end in the same suffix of the noun for a sort of assonance rule. Successively, articles start to be widespread and flexed. Articles seem to be the first part of speech related to the noun head to be acquired, probably for their proximity to the nouns, while later, attributive adjectives’ agreement, first for the number and then for gender, starts to be seen (Chini, 1995: 246). The agreement of predicative adjectives and past participles is the last to show up (Valentini, 1990; Chini, Ferraris, 2003). This confirms Greenberg’s Universal 31 (1966) which states that the agreement between noun and adjective is less marked than the one between noun and verb. Clitic pronouns appear even later in the IL of learners. Concluding, from the research done on the acquisition of the NP in Italian as a second language, it seems that gender errors are more frequent than those in number.

It should also be taken into consideration that the process of acquisition of gender, number and definiteness in the NP in Italian might also be influenced by extra-linguistic factors, i.e., the quantity and quality of the input, the age and gender of the learners, etc.

It can also be added that, according to Biazzini, Matteini (2010: 153), «in the management of sentence topics a Chinese learner of Italian faces two main tasks: learning both form and functions of new functional categories (e.g., articles and relative and clitic pronouns) and learning how to use L1-L2 common features (e.g., null subjects) according to the textual organizing principles of the target language».

3. THE RESEARCH: DESCRIPTION AND METHODOLOGY

3.1. *The learners’ profile*

Knowing precisely the sociolinguistic background of the participant in the study is fundamental for the study to be complete (Corino, Marengo, 2009). As mentioned above, the study takes into consideration the spoken recordings of 21 students. As far as known languages goes, all 21 of the students spoke Chinese mandarin and 11 of them could also speak the dialect¹³ of their hometown. Regarding foreign language competence, all of the 21 students could speak English, 4 of them spoke a second or third additional foreign language (1 spoke French, 1 Japanese, 1 Korean and 1 spoke French and a little

¹³ Some of the so-called “dialects” in China are spoken by millions of people and could almost be considered languages.

Norwegian). 3 students had studied abroad in English-speaking countries for a brief period before. In regards to their motivations for studying Italian as a major, while most of the students have an internal motivation to learn foreign languages (18 among 21), for 9 of them, Italian was not their first choice, and only 12 students stated that they were interested in Italian culture, of which 6 plan to continue their studies in Italy after graduation, but for a major related to economics or public communication and only 4 students hope to find a job related to Italian language and culture. In general, the motivation to study Italian is extrinsic.

In the period when the present research was conducted, the learners were in their first and second years of study and had 14 Italian lessons (each of 45 minutes) per week, among which, 10-12 were vocabulary lessons and 2 conversation lessons taught by Chinese teachers. From the second semester of the second year, the learners had 2 cultural lessons per week taught by Italian teachers. Learning activities mainly took place in a formal context in the classroom under the guidance of teachers, and the TL input was mainly from teachers (2 NNS and 1 NS), textbooks, recordings and classmates. The learners had very little exposure to native-speaker language.

Generally speaking, the learners' samples have certain foreign language learning aptitude (Dörnyei, Skehan, 2003: 597)¹⁴ and learning motivation, but their exposure to the TL is very low.

Referring to the three macro-phases of the development of the IL, the learners can be considered to be in the post-basic phase at the time of the research. In fact, they have passed the lexical stage and their IL shows gradually the specific features of the TL grammar, its categories and its morphosyntactic rules, with a finite utterance organization, corresponding in Jordens' words (2008) to the functional stage or finite linking stage.

3.2. *The input*

It is agreed by Instructed Second Language Acquisition (ISLA) scholars that formal class instruction can have an impact on second language learning but the debate is on how and how much influence it actually has (Ellis R., 2005; Loewen, 2015; Rebuschat, 2013; VanPatten, Williams, 2015; *inter alia*). Moreover, explicit declarative knowledge is not always necessarily followed by implicit knowledge: for example, a learner might know the rule but not apply it in a spontaneous, real-time oral interaction (DeKeyser, 2017; VanPatten, 2017; *inter alia*). For the population in this study, the exploration of Italian noun phrases occurred during the students' initial semester at university, spanning from September 2020 to January 2021. The structured learning sequence included: 1. understanding noun gender, 2. mastering indefinite articles, 3. grasping the usage of definite articles with singular nouns, 4. comprehending adjective gender agreement, 5. learning noun plurality, 6. understanding plural definite articles, 7. grasping plural adjective agreement, 8. mastering rules for modifying endings of masculine and feminine nouns based on natural gender (i.e., *commesso/commessa, direttore/direttrice*, and so on) 9. mastering demonstrative adjectives, 10. understanding qualitative adjectives, 11. grasping possessive adjectives, and understanding their usage in conjunction with definite articles. Grammar rules were directly introduced in Chinese within the textbook, supplemented by classroom explanations and practical examples. It should be noted that in the initial phase of studying Italian, examples were provided in the form of single phrases without context and

¹⁴ All students admitted in foreign language majors at SISU need to pass an exam of English language organized by the university in the first week of their studies. This is to ensure that they do not have any language learning disabilities and to put them in a suitable English language course.

grammatical rules were practised in a repetitive way. Hereafter are some examples taken from the textbook with explanation translated in English:

Lesson 6 - Grammar:

1. The number of definite articles. The definite article not only agrees in gender with the noun that follows it, but also matches in number, so the definite article also has singular and plural forms. The plural form is “i, gli, le”.
2. The number of adjectives:
 Adjectives ending in *-o* have four endings: *-o, -a, -i, -e*. Es. *l'asciugamano asciutto, gli asciugamani asciutti, la sciarpa sciupata, le scarpe sciupate*.
 Adjectives ending in *-e* have only two endings: *-e* and *-i*. Es. *il ragazzo forte, i ragazzi forti, la ragazza forte, le ragazze forti*.
 For adjectives ending in *-co*, if the stress falls on the penultimate syllable, the masculine plural ends in **-chi** and the feminine plural ends in *-che*. Es. *il fiore bianco, i fiori bianchi, la tazza bianca, le tazze bianche*.

Exercises:

1. Transform in plural the following nouns: il mese, la casa, lo zio, l'autobus, l'università, il signore, la chiave, la scuola, l'anno, l'operaia, il balcone, lo scudo, la zucca.
2. Transform the following words according to the example given “un bravo studente-bravi studenti”: una bella ragazza, un bambino forte, un giovane operaio, un tavolo lungo, uno zio serio.

By the end of January 2021, students had completed the study of the concordance of components in noun phrases, but the input consists almost entirely in explicit explanation and single exercises. From the second semester, the classroom input was changed from single phrases into short texts and dialogues, while grammatical emphasis was no longer put on the noun phrases, but on verb tenses and the use of articles.

3.3. Text typology

Part of the recordings (dating from Mar.5, 2021 to Dec.17, 2021) are oral productions which required the learner to undertake a structured task, e.g. to describe pictures that had been provided or to tell a story after viewing some images. There were in total 23 tasks undertaken in the duration of the two semesters. For each of the 11 tasks in the first semester, at least 2 images were provided, while in the second semester, for the 12 tasks some guided speaking lines and sample expressions were given to facilitate the student. The advantage in this case is that data are more easily compared and it could also help with the selection of verbal tasks that would lead the learners to use some specific morphosyntactic expressions to accomplish the exercise. On the other hand, the downside is that a structured task could prevent the production of more natural speech.

Not all students completed all 23 recording tasks due to some students missing part of the class. For the details of the recording including topics, student names, recording dates, recording duration and hours of Italian lessons attended by the students before each recording, see Appendix 1 and 2.

In addition to these recordings, the sample also includes live recordings of two final oral exams, recorded at the end of the first year and the end of the second year. The oral

exam is a one-on-one conversation between teachers and students. In the first exam, students were asked to describe a picture and answer some questions related to the details of the picture. In the exam for the second year, students were asked to express their opinion on a given topic and demonstrate their point of view. For the topic and duration of the recordings, see Appendix 3.

3.4. *The corpus*

The linguistic homogeneity of the learners, the absence of the TL context of immersion and the structured format are the main aspects distinguishing this corpus from those mentioned previously. The primary aim in the creation of this corpus was to carry out a longitudinal study to monitor how the IL of these learners in this specific context can change over time and successively to try and extract possible pedagogical implications.

As mentioned above, the corpus used in this study comes from the recordings of spoken productions of 21 Chinese students made between 2020 and 2021. The corpus has been divided into three sub corpora: the first one gathers all the recordings of spoken production tasks prepared by the learners over the second and third semesters and has a total of 70,975 tokens; the second comprises the recordings from the first year spoken exam, with a total of 5,227 tokens; while the third includes the recordings from the second year spoken exam, with a total of 4,804 tokens. The teacher's/interviewer's speech was removed from the transcript of the two exams.

To avoid the so-called “comparative fallacy” (Bley-Vroman, 1983), i.e., to compare the interlanguage (IL) of the learners to that of a supposed native speaker, the learner language in the three sub corpora was analysed cross-sectionally and longitudinally. Part of Speech (PoS) tagging was created to describe the IL of the students, as a mere observation of the phenomenon of the language of the students during the period when the recordings were collected, not as a way to signal the errors compared to the target language (TL). In fact, error-tagging with the purpose of seeing the difference between the IL and TL reveals even bigger theoretical issues. First, it contradicts the concept of IL as given by Selinker (1972), but they ultimately also concern the definition of TL itself. Nevertheless, at the same time the comparison with the L1 was only kept in mind for its pedagogical interest and possible influence on the syllabus and design of the materials used as well as the preferred teaching methodologies.

3.5. *Corpus tagging*

As mentioned earlier, the three corpora were analysed using a Part of Speech (PoS) annotation system (Tono, 2003) to solely describe the tokens in a systematic, explicit and transparent way, therefore making the study reliable and allowing the researchers to select and compare the data objectively. The use of the tagging gave the authors the required insight into the corpora, and the tool was chosen so as to avoid the comparative fallacy or “closeness fallacy” as coined by Klein and Perdue (1992). This opposed the comparative fallacy or, as in Klein and Perdue (1992), the “closeness fallacy” flaw, with the tagging being the tool enabling proper insight into the corpora, thereby allowing the researchers to fulfil their aims.

The software used for the analysis left the possibility of open-ended tagging, which meant that it could be open to all the variables of the speech context. Insufficient evidence to assign a tag might lead to ambiguous tags thus causing relevant problems in terms of reliability and objectiveness in the research (Milton, Chowdhury, 1994). In order to avoid

any interpretation on the part of the researchers, for the purpose of this study the tags used morphosyntactic descriptors. The descriptors referring to the NP could be linked to the morphosyntactic categories found in the noun phrase in Italian, such as “noun masculine/singular”, “determinative article feminine plural”, etc. The reference was taken, and later adapted by the authors, from Nuzzo (2010: 229) where the scholar proposes three different tables to guide in the observation of the learners’ IL: one for the noun phrase, the second for the verbal phrase and the last concerning the aspects of syntax and textuality. For the purposes of this study the first table on the NP was chosen and adapted to better fit the aim of the research. The adaptation consisted of breaking apart the questions in order to observe one aspect at the time; i.e., the question “are nouns and adjectives flexed by number and gender?” was divided into the following questions/tags: “is the noun flexed by number?”, “is the noun flexed by gender?”, “is the noun flexed by both number and gender?”, “is the adjective flexed by number?”, “is the adjective flexed by gender?”, “is the adjective flexed by both number and gender?”. Using this procedure the tagging became very detailed and, as a consequence, the researchers were able in the majority of cases to assign only one tag for each token, avoiding ambiguity. For the tokens that could be assigned multiple tags, the researchers created a specific tag called “more than one”, and the NP tagged in this way were later analysed manually one by one, breaking down each type of tag and reporting them in a spreadsheet chart to look for trends. In the three corpora objects of this study, the question of non-existent forms, as far as the TL was concerned, was irrelevant as no such case was found. As for the matter of deviant forms, the researchers used a specific tag, “deviant form”. Those tokens were successively analysed manually one by one by the researchers using the procedure of normalization: attributing divergent forms to existing plausible forms, therefore allowing a categorization of them which brought a deeper understanding of the form used by learners in their IL. Thus the annotation can be seen on a double level: the “source” one and the “tendential” one (Andorno, Rastelli, 2009: 60-61). The source one is that found in the corpus, while the tendential one is the form the token “tends to” in the TL. Doing this, it is less likely for the researcher’s point of view to pollute the data with his/her interpretation, since the annotation does not involve for the same categories the different forms in IL and TL, but the same form (the one in IL) according to different perspectives.

Moreover, the software used for this study enabled visualisation of each token in context, thereby allowing the researcher to analyse the chunk where it was found. *Chunk* is defined as a textual unit composed of adjacent words morphologically analysed, where words have the property of being correlated for dependency chains (such as determination, quantification, etc.) and can be identifiable without ambiguity.

4. DATA ANALYSIS

4.1. *The general description of NPs in three corpora*

The corpus contains a total of 81,006 tokens and 17,723 NPs.

Sub corpus 1 includes the texts of prepared oral productions and consists of 16,407 NP, representing 23.1% of the total number of tokens. Of all NPs, 14,303 are correct, 45 are self-corrections made by students, and 2,059 are errors. Among the NPs in error, 81 have more than one error, which constitutes 0.039% of the total number of errored NPs. The total number of errors in noun phrases is 2,112.

Sub corpus 2 includes recordings of the first-year final exam and consists of 871 NPs, representing 16.7% of the total number of tokens. Within this corpus, there are 596 correct NPs, 46 corrections made by students (5.28% of the total), 54 repetitions (6.19%

of the total), and 175 with errors, 7 of which have more than one error, equating to 0.013% of the total number of NPs with errors. The total number of NP errors is 182.

Sub corpus 3 encompasses the recordings from the second-year final exam and contains a total of 805 NPs, representing 16.8% of the total number of tokens. Among these, there are 586 correct NPs, 41 corrections made by students (5.09% of the total), 31 repetitions (3.85% of the total), and 219 with errors, 7 of which have more than one error, equalling 0.032% of the total number of NP errors. The total number of errors in NPs is 227.

The frequencies of both corrections and errors in sub corpus 1 are significantly lower than those in sub corpus 2 and 3, which is consistent with the expected qualities of prepared texts versus improvised texts.

It should be noted that the repetition rate decreased by 62% from the 1st to the 2nd year's exam. This phenomenon is also addressed by Andorno (2011: 46) in a study that compares two groups of Chinese students: one in a natural environment and one in a structured context, which shows that in the latter, there are more repetitions and corrections and the discourse is more fragmented.

4.2. Statistical analysis

In this paragraph each of the main aspects found will be considered individually.

4.2.1. Error rate related to each PoS in NP

To obtain the proportion of NP errors related to a particular part of speech, the total number of errors for each category in the three sub corpora is compared to the total number for each corresponding part of speech. The error rate of each PoS is reported in Table 2.

Table 2: *Error rate of each PoS in NP in three sub corpora*

PoS	Sub corpus 1			Sub corpus 2			Sub corpus 3		
	n. token	n. err.	% err.	n. token	n. err.	% err.	n. token	n. err.	% err.
noun	16407	427	2.60%	871	46	5.28%	805	42	5.22%
adjective	4975	268	5.39%	326	20	6.13%	279	20	7.17%
article	7754	474	6.11%	554	35	6.32%	459	50	10.89%
demonstrative	415	9	2.17%	48	1	2.08%	57	2	3.51%
possessive	2021	39	3.82%	81	5	6.17%	82	4	4.88%
infinite	776	5	0.64%	41	3	7.32%	67	0	0.00%
preposition	9842	576	5.85%	563	46	8.17%	550	26	4.73%

Comparing the data in the prepared recordings with that from the two exams, it was observed that the error rate in the exams is generally higher than that in the recordings.

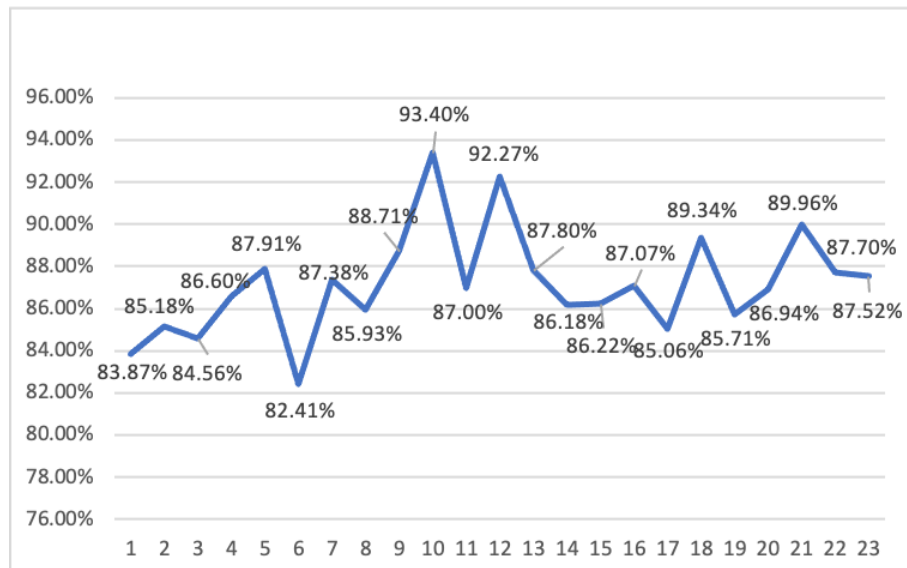
This is especially evident from the fact that the error rates related to the use and form of nouns in the exams were almost twice that of the recordings. This discrepancy could be attributed to factors such as stress or anxiety caused by the exam situation, thus we see a difference in the output derived from the use of either the procedural or the declarative memory (Paradis, 2004).

By comparing the data from the two exams and removing data on possessives, infinitives, and demonstratives that are too few to be representative, it was found that errors related to nouns and adjectives remained consistent. Errors related to prepositions significantly decreased in the second year of study, while those related to articles show a significant increase. From this, it can be inferred that during the exams students make use of those chunks of languages, especially those containing prepositions, that they are sure of, as a strategy to make fewer mistakes. This could also be a sign that shows that in the IL of the students' sample, the fixed usages of specific prepositions in the Italian language could be in the process of being acquired. Detailed analysis and examples will follow.

4.2.2. Correct rate of NP in prepared texts

After calculating the average correct rate of NP in each assignment of the 23 students and arranging them chronologically, it was observed that the correct rate fluctuated initially before increasing during the 11 recordings from March to June 2021, peaking at the end of the first year (93.40%). However, since September 2021, the correct rate has fallen back and then fluctuated within a certain range (min. 85.06%, max. 89.96%).

Table 1. *Average correct rate in each assignment*



Successively, the 23 recordings were divided into two parts based on the date of recording: recordings 1 to 11 were produced in the second semester of the first year and recordings 12-23 in the first semester of the second year. The standard deviation of the average correct rate of students for both groups was calculated and the data of the first group has a greater standard deviation (0.029) than that of the second group (0.020). Next, using the hours of lessons taken before each recording date as the independent variable and the average correct rate as the dependent variable, Pearson correlation coefficients

were calculated. For the first 11 recordings, the correlation coefficient (r value) was 0.64, indicating a strong correlation between the increase in studying hours and the correct rate of noun phrases from March to June 2021. On the other hand, for the next 12 recordings, the correlation coefficient was -0.14, indicating a very weak correlation. This means that, after a beginning phase of evident improvement in NP correctness, from September to December 2021, the correct rate of NP did not increase despite the increase in studying hours. This could be due to limited exposure to TL inputs and a restricted study duration. Additionally, students had not received lessons from a native teacher throughout their initial three semesters of university, which may have caused the persistence of misuse in nouns and adjectives.

4.2.3. *Study hours and number of errors related to each PoS*

Moreover, Pearson correlation coefficients were computed between study hours before each recording as the independent variable and the total number of four types of errors related to nouns, adjectives, articles, and prepositions as the dependent variables. The results demonstrate a strong positive correlation ($r = 0.67$) between the increase in preposition errors and the increase in learning time, which differed significantly with the decrease in preposition errors between the two exams. The reason might be that some preposition-noun combinations were memorized by the students and in exams they tended to use only a limited number of prepositional phrases familiar to them in order to avoid mistakes. In fact, after comparing the frequency of prepositions respectively in three sub corpora, it was found that the students used the prepositions more frequently in prepared texts than in exams. The total frequency of prepositions in sub corpus 1 consists of 12% of total tokens, while in sub corpus 2 and 3, the percentage was respectively 7.9% and 9.3%. Moreover, comparing the recordings of the two exams, it was observed that fixed structures like *per esempio* ('for example'), *di solito* ('usually') and some structures learned at the pre-basic phase of study, such as *a Shanghai* ('in Shanghai'), *all'estero* ('abroad'), *di* + name of a person (of sb.) were repeatedly used, thus contributing to the decrease in error occurrence in exams.

The increase in noun errors and the increase in article errors also have a positive correlation with learning time, but less strongly, with r values of 0.50 and 0.28, respectively. However, the number of adjective errors does not show a significant correlation with learning time ($r = 0.07$). Other studies (Chini, 1995; Bartning, 2000; Dewaele, Véronique, 2000; Granfeldt, 2005) have shown that the agreement between nouns and adjectives is tendentially learnt later than that between nouns and articles.

4.2.4. *Agreement*

In the recordings of prepared texts, there are a total of 203 incorrect noun phrases with errors related to gender and number agreement. Results from word frequency statistics are as follows: there were 138 nouns in total, out of which 49 ended in *-a* and appeared a total of 79 times, accounting for 38.9%; 41 ended in *-o* and appeared 58 times, accounting for 28.6%; 39 ended in *-e* and appeared 56 times, accounting for 27.6%; and 10 ended in other letters, accounting for 4.9%. Among nouns ending in *-e*, 14 ended in *-ione* and appeared 20 times, accounting for 35.7% of the total frequency of nouns ending in *-e*.

The most frequent word with agreement errors is *persona* ('person'), which appeared 9 times in total, all in plural form *persone* ('people'), possibly showing a result of transfer

from English, the first LS learned by students, in an attempt to follow the same rule of the words *student* and *professor* in English that become *studente* and *professore* in Italian. In these 9 sintagma, 6 appeared with a masculine singular flexion of the adjective, twice with a masculine plural flexion of the adjective, and once with a feminine singular form of the definite article. It is interesting that in the recordings of the two exams, the frequency of the word *persona* was respectively 11 and 25, of which only 2 with agreement errors:

1. **troppo persone* (correct form: *troppe persone*)
too many people
2. **diciassette persona* (correct form: *diciassette persone*)
seventeen people

The second NP was followed by the correction from the student herself.

3. *diciassette persone*

This could show that, after an initial period of confusion, most students have surpassed the difficulty of attributing the female gender to a word that semantically refers to both genders.

Other frequent errors include:

- Noun forms ending in *-e*, either singular or plural, are commonly considered as masculine singular.
 4. **l'altro parte*, **un parte*, **il maggior parte* (correct form: *l'altraa parte*, *unaa parte*, *la maggior parte*)
another part, a part, the most part
 5. **dal boutique* (correct form: *dallaa boutique*)
from the boutique
 6. **secondo classe* (correct form: *secondaa classe*)
second class
 7. **del bevande* (correct form: *dellee bevande*)
some drinks
- Errors are frequently made in abbreviations, compound words, and some masculine nouns ending in *-a*.
 8. **il pallacanestro*, **il pallavolo* (correct form: *laa pallacanestro*, *laa pallavolo*)
(the) basketball, (the) volleyball
 9. **il mio foto*, **un foto* (correct form: *laa mia foto*, *unaa foto*)
my photo, a photo
 10. **il metro* (correct form: *laa metro*)
the metro
 11. **la problema*, **la programma* (correct form: *ilo problema*, *ilo programma*)
the problem, the program
 12. **la media* (correct form: *ii media*)
the media

- When nouns are attributed with the correct gender and number, students tend to use adjectives with the same ending letters as the nouns being modified, resulting in errors such as
 13. **il clima è perfetta* (correct form: *il clima è perfetta*)
the weather is perfect
 14. **le cose importante* (correct form: *le cose importanti*)
(the) important things
 15. **la cravatta verda* (correct form: *la cravatta verde*)
the green tie
 16. **le feste importante* (correct form: *le feste importanti*)
(the) important festivities
 17. **le persone sono felice* (correct form: *le persone sono felici*)
the people are happy

- Fixed structures, such as *pieno di* and *vicino a*, tend to remain invariable. Considering the presence of a preposition in these structures, the students might have taken them as a prepositional phrase and thus ignored the necessity of flexion.
 18. **la stanza pieno di oggetti, *la strada pieno di acqua* (correct form: *la stanza piena di oggetti, la strada piena di acqua*)
the room full of objects, the street full of water
 19. **la bambina vicino a mia madre, *una scuola vicino a casa* (correct form: *la bambina vicina a mia madre, una scuola vicina a casa*)
the girl near my mother, the school near my home

4.2.5. Word order

Among the 84 errors tagged as “other errors” in sub corpus 1, 40 were related to word order, of which 18 had a modifier that preceded the relevant nouns, 18 with a modifier after the nouns, 2 related to the position of adverbs and 2 considering the use of *tutto* with determinative articles. Errors in the order of words might be attributed to L1 language transfer: in fact in Mandarin Chinese, all the modifiers precede the relevant nouns (Chan, 2014). But it is interesting to notice that errors with the postposition of modifiers occurred mostly within the period when the students had just learnt the use of possessives. The modifiers in question were concentrated in several adjectives: *suo* (‘his, her’), *loro* (‘their’), *altro* (‘of another person’), *proprio* (‘of oneself’) and ordinal numbers.

20. **nella casa sua* (correct form: *nella sua casa*)
in his house
21. **gli abiti propri* (correct form: *i propri abiti*)
his own clothes
22. **al piano terzo* (correct form: *al terzo piano*)
on the third floor
23. **in campi altri* (correct form: *in altri campi*)
in other fields

These errors were frequent in that period, showing that students were applying the canonical noun-modifier order of TL to the newly-learned modifiers, possessives in this case. After 500 hours of lessons, errors of word order showed up randomly.

4.2.6. *Improper use of articles*

60.2% of the improper use of articles are omissions, with the frequency of 203 out of 337. Nonetheless it should be noted that 109 on 203, so almost 54% of the article omissions occur together with a preposition, which could mean that the issue is not when and where to use the article zero, but when to use the simple preposition or the preposition with the article. If so, it would mean that the matter is about the definitiveness of the NP. Other types of errors in the articles are connected with definitiveness (32), overuse (96) and overlapping (6). The occurrence of omissions increased with the increase in studying hours ($r = 0.499$), while the occurrence of the other three types of errors are not correlated with the studying hours (definiteness, $r = -0.060$; overuse, $r = -0.088$; overlapping, $r = -0.246$). 109 of the omissions occurred when a preposition was used, 14 with a possessive, 36 in a coordinate structure comprising more than two nouns. The omissions with a possessive should be attributed to LS transfer, since in English there is no article before a possessive. This kind of error is always observed in class and in students' written texts in the beginning phase of study. The omissions in a coordinate structure were observed mostly in the recordings of the argumentative texts and might be influenced by the omissions of articles in some informal oral/written texts, like interviews and posts on the Internet, studied in class. Similar results about spoken learners' texts can be found in Chiapedi (2010: 72).

4.2.7. *Improper use of nouns and adjectives*

Errors of improper use are found to be the majority of errors in nouns and adjectives in prepared texts with a percentage respectively of 67.7% of all noun errors and 56.8% of all adjective errors. The learners, albeit with the help of dictionaries and translating tools, still produced a large amount of errors in selecting the right nouns and adjectives. This shows a persistent difficulty in understanding the semantic field which a specific noun or adjective belongs to. After analysing those errors manually one by one, a strong correlation between the misused words and their corresponding translation in Chinese mandarin or in English was found, showing the presence of L1/LS transfer or crosslinguistic influence.

4.2.7.1. *Transfer from L1*

In Chinese mandarin there are no flexion forms of nouns and adjectives and in many cases the same word can be attributed to more than one PoS, especially for nouns and verbs. In many cases, a noun and a verb with the same form assume different meanings, but there are still numerous cases in which they share the same meaning, or nearly, but emphasize different aspects in different contexts. Usually, when the word is used as a noun, it implies the concept or the content of the thing to which the word refers, and when used as a verb, the emphasis lies on the action itself. In fact, in *Xian Dai Han Yu Ci Dian* (ed. 1996) there are 767 pairs of words with identical forms and 22% of them consist

of a noun and a verb that are close in meaning but related to different semantic fields. The 110 pairs of words collected in the dictionary account for 45.4% of the total number of words that may be used as both a noun and a verb (Su, 2000: 31). This could be an explanation of the frequent misuse of noun forms instead of the corresponding verb in IL of the Chinese students. Following are some examples:

24. *Ci vado spesso per *lavoro (lavorare) e leggere.*
I go there often for work and reading.
25. *Hanno deciso di *divorzio (divorziare).*
They decided to divorce.
26. *... lo spazio per *studio (studiare)*
... the space for study
27. *Vado a *sport... (fare sport)*
I'm going to do some sport...
28. *Vado a *shopping... (fare shopping)*
I'm going shopping...
29. *Insomma, *viaggio (viaggiare) in treno è comodo.*
In short, traveling by train is comfortable.
30. *Lavoro (lavorare) felicemente è la cosa più importante.*
Working happily is the most important thing.

Transfer from L1 may also be seen in the choice of some particular adjectives, like the use of *lontano* ('far') instead of *lungo* ('long') when describing a distance and the use of *caro* ('expensive') instead of *alto* ('high') when modifying the word *prezzo* ('price'). These errors have been repeatedly found in sub corpus 1 and may be due to the validity of both adjectives in Chinese when describing the subjects in question.

Some errors concerning the semantic fields of synonyms may also be attributed to the lack of natural input with context of use and the cultural differences. For example:

31. *un negozio di vestiti (abbigliamento)*
a clothes shop
32. *ha comprato qualche abbigliamento (vestito)*
she bought some clothes
33. *la camera della festa (sala, stanza)*
the party room
34. *va dal *barbiere (parrucchiere).*
She's going to the hairdresser.
35. *ho chiesto *un nonno (anziano)*
I asked an old man.

In cases (34) and (35), cultural difference plays a key role in the choice of the noun. In hair salons in China there is no distinction between a *barbiere* ('barber, gentlemen's hairdresser') and a *parrucchiere* ('hairdresser'), with both called 理发师 *lǐfà shī* 'master in hair cut'. In the second case, a well-educated person would call any old man with a similar age of their grandfather 爷爷 *yéyé* 'grandpa' to show respect, but in Italian the word *nonno* ('grandpa') refers only to the father of one's father/mother.

4.2.7.2. *Transfer from LS*

Forty-two cases of improper use of nouns could be attributed to the negative transfer from English, since most of the words used by the learners have a morphologically similar correspondent word in English, which in English belongs to the same semantic field in which the correct Italian word should be located. Two typical errors in this category were the use of *pittura* ('picture') instead of *foto* ('photo') or *immagine* ('image') to indicate a picture and the use of *classe* ('class') instead of *lezione* ('lesson') when referring to a language class, very frequent in the recordings of the first assignments. Similar mistakes can be found in various places when learners used *appareanza* to indicate the appearance, *applicazione di Master* to indicate the application for a Master course, *preiscrizione all'università* for the pre-enrolment in university and *graduazione* for graduation. The similarity in morphology between some words in English and in Italian could facilitate the acquisition of new lexicals in the TL, but, without input from natural contexts, the learners had to activate their knowledge and experience obtained in previous FL learning and thus formulated phrases with the words that sounded similar but in fact should be used in a different context.

Transfers from English in improper use of adjectives are fewer than that of noun errors, with only 14 cases, but the errors were not a mere transformation of morphologically similar words. Following are some examples.

36. *Tutto sommato viaggiare in treno è rilassante, chiacchierare con i *suoi (propri) amici e ammirare i paesaggi.*
All in all, traveling by train is relaxing, chatting with your friends and admiring the landscape.
37. *In una grande città moderna come Shanghai la gente è occupata cammi- camminano alla strada con i *suoi (propri) amici oppure il *suo (proprio) cellulare e ci sono molti forestieri che non sanno i percorsi ancora.*
In a big, modern city like Shanghai, people are busy walking - they walk the street with their friends or their mobile phones, and there are many foreigners who don't know the routes yet.
38. *I giovani appena laureati raramente riescono a trovare un lavoro senza stress ed è adatto alle *sue (proprie) esigenze.*
Young college graduates rarely find a job that is stress-free and suited to their needs.

In the above phrases the possessive *suo*, which corresponds to *his/her* in English, was used instead of *proprio* to refer to the agent of the verb, while in Italian, when the agent does not refer to a determinate person, *suo* should not be used. This phenomenon of transfer might imply that students had difficulties in understanding the semantic fields which *suo* and *proprio* belong to.

About cross linguistic transfer, a similar situation, where existing forms in the TL are used but with different meanings can also be found in the corpus study run in Perugia (Atzori, Chiapedi, Spina, 2009: 103). Nevertheless, cross linguistic influence should not change the acquisitional sequences, but only accelerate or slow the process of SLA (Chini, 2010: 35-36)¹⁵.

Related to this aspect, Hoey (2009: 1) says:

¹⁵ For an insight on crosslinguistic influence on the acquisition of the null-subject in Chinese students, see Cherici (2021).

This process of subconsciously noticing is referred to as lexical priming. Noticing all these things is what makes it possible for a speaker to use the right phrase in the right context at the right time. Without realizing what we are doing, we all reproduce in our own speech and writing the language we have heard or read before. We use the words and phrases in the contexts in which we have heard them used, with the meanings we have subconsciously identified as belonging to them and employing the same grammar. This is how native speakers are able to be fluent and because the things they say are subconsciously influenced by what everyone has previously said to them, it also explains why they almost always sound natural. [...] The differences in practice between a native speaker and a non-native speaker are twofold. Firstly, a non-native speaker is typically exposed to less language and to a narrower range of language, and, secondly, the non-native speaker has previously been primed for another language, which initially affects the way he or she is primed in English. Native speakers have acquired a large corpus of examples of the words of English in their typical contexts, and from this they learn how the words are used. By contrast, non-native speakers have typically heard (or read) relatively few examples of even the more common words in natural use and have therefore had less opportunity to learn the way these words typically occur.

4.2.8. *Repetitions and corrections*

In the recordings of the two oral exams there was a relatively high percentage of correction and repetition in the total amount of NPs: 5.28% that of correction and 6.19% of repetition in exam 1 and 5.09% of correction and 3.85% of repetition in exam 2. The total percentage of correction and repetition was lower in exam 2 compared to exam 1, showing an improvement in production fluency after one year of study.

It was observed that, in exam 1 most repetition cases occurred at the beginning of a phrase with a repetition of the subject (16 cases) or at the end of a phrase (10 cases), probably aimed to gain more time for a better structuring of the following phrase. This kind of repetition is reduced drastically in exam 2, with only 2 cases at the beginning of a phrase and 1 case at the end. In exam 2, most repetitions occurred when the student wanted to add a modifier afterwards, like *la situazione la situazione della società, la casa la casa mia* and in some cases the repeated noun was followed by a relative clause to add more information, like *un lavoro un lavoro che può guadagnare* or by a prepositional phrase, like *il più importante cosa cosa da ottenere*.

In the cases of correction, in exam 1 the substitution of flexural elements were the most frequent (23 cases, 50%), among which 18 related to the flexion of gender (9 of adjectives, 6 of articles, 3 with nouns) and 5 related to the flexion of number (4 of nouns, 1 of article). A similar situation pertained in exam 2, where 14 cases (34%) were related to flexural elements, 11 of gender and 3 of number. Another frequent type of correction was the completion of a partially pronounced noun, like *lasa (lasagne), abba (abbigliamento), le pa (i pantaloni), la mo (la moglie)*. This kind of correction occurred both in exam 1 and in exam 2, with a similar frequency of 11 cases versus 10. The nouns repeated were usually longer and complicated in morphology, showing that the morphological complexity would influence the fluency of oral production in TL. Other corrections were related to the correction of pronunciation and the insertion of prepositions and pronouns, but with very few cases, not enough to be representative. Lexical substitution was not found in either exam.

The huge difference seen in the corpus which contains the prepared oral production of the students compared to the other two sub corpora of the exams can be attributed to the fact that for learners it is difficult to pay attention to language forms and to content at the same time, leading to shifting attention to content over form (Van Patten, 1990; Schmidt, 2000). Another possible explanation could be the use of different types of memory. In fact, while structured teaching/learning in class promotes conscious learning of the rules and communicative practice routines supported by declarative memory, Ellis R. (1994a: 86) proposes a model where implicit and explicit knowledge combine with controlled and automatic processing. Acquisition is thus a continuous process among those leading to the automatization of the TL. But according to Paradis (2004), implicit and unconscious knowledge cannot be used in a controlled way. According to him (2008), the linguistic competence is acquired accidentally, stored implicitly and used automatically, supported by the procedural memory (Ullman, 2001; 2005). Practice could speed up the usage in reproducing the grammar rules but it doesn't necessarily mean that those rules are interiorized (Hulstijn, 2003). What has just been stated should be taken into deep consideration for the impact it could have on language teaching practices in the classroom and how the language course syllabus should be planned, also in the light of the Teachability Theory (Pienemann, 1984; 1998: 252).

5. CONCLUSIONS AND LIMITATIONS

The limit of the present research is that the corpus size is small and therefore, more research is needed to make more reliable generalizations.

Nonetheless, looking at the results of this study, the authors feel that similar conclusions to previous studies on acquisitional sequences in Italian as a second language can be drawn for Italian as a foreign language. In fact, it seems that the acquisitional sequences for learners in a structured context follow the same order as those of learners in a natural environment that do not undertake structured input. But it can be said that structured input can speed up the acquisition process since it favours noticing the morpho-syntactic and lexical-grammatical features of the TL and their production (Andorno, 2011: 47). In this sense the crucial role of input (Gass, Madden, 1985; White, 1987; VanPatten, 1990; Gass, 1997; Piske, Young-Scholten, 2009; Borro, Scolaro, 2021; *inter alia*), interaction (Long, 1996; 1980; 1981; 2020; *inter alia*) and output (Gass, 1997; Nuzzo, Grassi, 2016) must be acknowledged. The statistical analysis done on the prepared recordings, corpus number one, shows a strong correlation between class time and errors at least in those of the first part, while less so in the second part. This could mean that instructed learning could enhance students' development of language competence but input and interaction, especially with NNS, are fundamental to sustain/maintain the process of developing it. One crucial difference found between the prepared tasks recording and the two exam sessions is that the rate of mistakes on NPs is definitely much higher during the examinations. An improvement of almost 50% can also be seen between the first and second year exam recordings. As mentioned earlier, this data can show the influence of anxiety when performing a stressful task as a situation like the examination requires, but also the different types of memories working underneath the performance, declarative in the prepared recording and procedural in the case of the exams, as well as the fact that during the exam performance, learners focus more on content than on form. Moreover, the contrast between the first and second year recordings and the students' improvements could show that some morphosyntactic structures are in the process of being learned.

In addition, the context of use of lexicons in guided learning merits more attention, not only on the teaching of semantic fields and words used in the appropriate context, but also considering the cultural difference behind the choice of words.

Moreover, data-driven Learning (DDL) could also be very effective in teaching lexico-grammatical features (Boulton, Cobb, 2017; Boulton, Vyatkina, 2021; Forti, Spina, 2019; Forti, 2021, 2023; Spina, Forti, Grego Bolli, 2020; Zanca, 2018) assuming that the validity of this methodology is not merely on reaching the linguistic and grammatical competence in the TL as much as in the process of learning and in the acquisition of strategies and awareness on how to deal with language questions.

REFERENCES

- Andorno C., Rastelli S. (eds.) (2009), *Corpus di Italiano L2: tecnologie, metodi, spunti teorici*, Guerra Edizioni, Perugia.
- Andorno C. (2010), “Lo sviluppo della morfosintassi in studenti cinesi”, in Rastelli S.(ed.), *Italiano di cinesi, italiano per cinesi - Dalla prospettiva della didattica acquisizionale*, Guerra, Perugia, pp. 89-122.
- Andorno C. (2011), “La grammatica per l'apprendente di L2. Apprendenti guidati e spontanei a confronto”, in Corrà L., Paschetto W. (eds.), *Grammatica a scuola*, FrancoAngeli, Milano, pp. 36-48.
- Atzori L., Chiapedi N., Spina S. (2009), “Corpora di italiano L2: difficoltà di annotazione e trascrizione “allargata””, in Andorno C., Rastelli S. (eds.), *Corpus di Italiano L2: tecnologie, metodi, spunti teorici*, Guerra Edizioni, Perugia, pp. 93-110.
- Barni M., Gallina F. (2008), “Le parole degli stranieri: il LIPS, il primo lessico di frequenza dell'italiano parlato dagli stranieri”, in Barni M., Troncarelli D., Bagna C. (eds.), *Lessico e apprendimenti. Il ruolo del lessico nella linguistica educativa*, FrancoAngeli, Milano, pp. 143-156.
- Barni M., Gallina F. (2009), “Il corpus LIPS (Lessico dell'italiano parlato da stranieri): problemi di trattamento delle forme e di lemmatizzazione”, in Andorno C., Rastelli S. (eds.), *Corpus di Italiano L2: tecnologie, metodi, spunti teorici*, Guerra Edizioni, Perugia, pp. 139-151.
- Bartning I. (2000), “Gender agreement in L2 French: pre-advanced vs advanced learners”, in *Studia Linguistica*, 54, 2, pp. 225-237.
- Bernini G. (2003), “Come si imparano le parole. Osservazioni sull'acquisizione del lessico in italiano L2”, in *Bollettino Itals*, 1/2, pp. 23-47.
- Berretta M. (1990), “Morfologia in italiano lingua seconda”, in Banfi E., Codin P. (eds.), *Storia dell'italiano e forme dell'italianizzazione*. Atti del XXIII 131 Congresso internazionale di studi: Trento-Rovereto 18-20 maggio 1989, Bulzoni, Roma, pp. 181-201.
- Biazzi M., Matteini I. (2010), “Referential and topic movement in Chinese learners of Italian: a longitudinal account”, in Chini M. (ed.), *Topic, struttura dell'informazione e acquisizione linguistica*, FrancoAngeli Editore, Milano, pp. 137-158.
- Biber D., Johansson S., Leech G., Conrad S., Finegan E. (1999), *Longman grammar of spoken and written English*, Longman, New York.
- Bley-Vroman R. (1983), “The comparative fallacy in interlanguage studies: the case of systematicity”, in *Language Learning*, 33, pp. 1-17.
- Borro I., Scolaro S. (2021), “Optimal and appropriate input in a second language: The potential of (modified-)elaborated input in distance and classroom learning”, in *E-*

- Journ.ALL, Euro-American Journal of Applied Linguistics and Languages Special Issue*, 8, 2, pp. 53-69: <http://doi.org/10.21283/2376905X.14.246>.
- Boulton A., Cobb T. (2017), "Corpus use in language learning: A meta-analysis", in *Language Learning*, 67, 2, pp. 348-393.
- Boulton A., Vyatkina N. (2021), "Thirty years of data-driven learning: Taking stock and charting new directions", in *Language Learning and Technology*, 25, 3, pp. 66-89.
- Brezina V., McEnery T. (2021), "Introduction to Corpus Linguistics", in Tracy-Ventura, N., Paquot M. (eds.), *The Routledge Handbook of Second Language Acquisition and Corpora*, Routledge, New York, pp. 11-22.
- Callies M., Götz S. (2015), "Learner Corpora in Language Testing and Assessment. Prospect and Challenges", in Callies M., Götz S. (eds.), *Learner Corpora in Language Testing and Assessment*, John Benjamins, Amsterdam-Philadelphia, pp. 1-9.
- Chan A. Y. W. (2004), "Noun phrases in Chinese and English: a study of English structural problems encountered by Chinese ESL students in Hong Kong", in *Language, Culture and Curriculum*, 17,1, pp. 33-47: <https://doi.org/10.1080/07908310408666680>.
- Cherici A. (2021), "The role of L1 and L2 in the acquisition of null subjects by Chinese learners of L3 Italian", in *International Journal of Multilingualism*, pp. 735-752: <https://doi.org/10.1080/14790718.2021.1977308>.
- Chiapedi N. (2010), "L'articolo italiano nell'interlingua di apprendenti sinofoni: problematiche acquisizionali e considerazioni glottodidattiche", in *Italiano LinguaDue*, 2, 2, pp. 53-74: <https://riviste.unimi.it/index.php/promoitals/article/view/822>.
- Chini M. (1995), *Genere grammaticale e acquisizione. Aspetti della morfologia nominale in italiano L2*, FrancoAngeli, Milano.
- Chini M., Ferraris S. (2003), "Morfologia del nome", in Giacalone Ramat A. (ed.) *Verso l'italiano. Percorsi e strategie di acquisizione*, Carocci, Roma, pp. 37-69.
- Chomsky N. (1959), "Review of "Verbal Behavior" by Skinner B.F.", in *Language*, 35, pp. 26-58.
- Chomsky N. (1981), *Riflessioni sul linguaggio*, Einaudi, Torino.
- Corbett G. (1991), *Gender*, Cambridge University Press, Cambridge.
- Corino E., Marelllo C. (2009), "Didattica con i corpora di italiano per stranieri", in *Italiano LinguaDue*, 1, 1, pp. 279-285: <https://riviste.unimi.it/index.php/promoitals/article/view/444>.
- Dardano M., Trifone P. (1995³), *Grammatica italiana con nozioni di linguistica*, Zanichelli, Bologna.
- Dewaele J., Véronique D. (2001), "Gender assignment and gender agreement in advanced French interlanguage: a cross sectional study", in *Bilingualism: Language and Cognition*, 4, 3, pp. 275-297.
- DeKeyser R. (2017), "Knowledge and skill in ISLA", in Loewen S., Sato M. (eds.), *The Routledge handbook of instructed second language acquisition*, Routledge, New York, pp. 15-32.
- Dörnyei Z., Skehan Z. (2003), "Individual Differences in Second Language Learning", in Doughty C. J., Long M. H. (eds.), *The Handbook of Second Language Acquisition*, Blackwell Publishing Ltd., Oxford, pp. 589-630: <https://doi.org/10.1002/9780470756492.ch18>.
- Ellis R. (1994a), "A theory of instructed second language acquisition", in Ellis N. C. (ed.), *Implicit and explicit learning of languages*, Academic Press, London, pp. 79-114.
- Ellis R. (1994b), *The study of second language acquisition*, Oxford University Press, Oxford.
- Ellis R. (2005), *Instructed second language acquisition: A literature review*, Ministry of Education, Wellington, New Zealand.

- Forti L. (2021), “Apprendere la grammatica attraverso il Data-driven learning”, in Giunchi P., Roccaforte M. (eds.), *La grammatica tra acquisizione e apprendimento. Un percorso verso la consapevolezza linguistica*, Carocci, Roma, pp. 100-113.
- Forti L. (2023), *Corpus use in Italian language pedagogy. Exploring the effects of data driven learning*, Routledge, New York.
- Forti L., Spina S. (2019), “Corpora for linguists vs. corpora for learners: Bridging the gap in Italian L2 learning and teaching”, in *ELLE - Educazione linguistica. Language education*, 8, 2, pp. 349-362.
- Gass S. (1997), *Input, interaction and the second language learner*, Lawrence Erlbaum Associates Publishers, Mahwah (NJ).
- Gass S., Madden C. (eds.) (1985), *Input in second language acquisition (Series on Issues in Second Language Research.)*, Newbury House, Rowley (Mass.).
- Gass S., Selinker L. (2001), *Second Language Acquisition. An Introductory Course*, Lawrence Erlbaum, Mahwah (NJ).
- Giacalone Ramat A. (ed.) (2003), *Verso l'italiano. Percorsi e strategie di acquisizione*, Carocci, Roma.
- Gilquin G. (2015), “From design to collection of learner corpora”, in Granger S., Gilquin G., Meunier F. (eds.), *The Cambridge Handbook of Learner Corpus Research*, Cambridge University Press, Cambridge, pp. 9-34:
<http://hdl.handle.net/2078.1/145509>.
- Granfeldt J. (2005), “The development of gender attribution and gender agreement in French: a comparison of bilingual first and second language learners”, in Dewaele J. (ed.), *Focus on French as a foreign language*, Multilingual Matters, Clevedon, pp. 164-190.
- Granger S. (2002), “A Bird’s-Eye View of Learner Corpus Research”, in Granger S., Hung J., Petch-Tyson S. (eds.), *Computer learner corpora, second language acquisition and foreign language teaching*, John Benjamins, Amsterdam-Philadelphia, pp. 3-33:
<https://doi.org/10.1075/llt.6.04gra>.
- Granger S. (2009), “The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation”, in Aijmer K. (ed.), *Corpora and Language Teaching*, John Benjamins, Amsterdam-Philadelphia, pp. 13-32.
- Granger S. (2012), “How to Use Foreign and Second Language Learner Corpora”, in Mackey A., Gass S. M. (eds.), *Research methods in second language acquisition. A practical guide*, Wiley-Blackwell, Oxford, Hoboken (NJ), pp. 7-29:
<https://doi.org/10.1002/9781444347340.ch2>.
- Greenberg J. H. (ed.) (1963), *Universals of language: report of a conference held at Dobbs Ferry, New York, April 13-15, 1961*, MIT Press, Cambridge (Mass.).
- Holmes V. M., Segui J. (2006), “Assigning grammatical gender during word production”, in *Journal of Psycholinguistic Research*, 35, 2, pp. 5-30.
- Hulstijn J. H. (2003), *Incidental and Implicit Learning*, John Benjamins, Amsterdam-Philadelphia.
- Iurato A. (2022), “Learner Corpus Research meets Chinese as a Second Language Acquisition: Achievements and Challenges”, in *Annali di Ca’ Foscari, Serie Orientale* 58, 1: [1-34], pp. 709-742.
- Klein W., Perdue C. (1992), “Utterance structure”, in Perdue C. (ed.), *Adult Language Acquisition: cross-linguistic perspectives*, vol. 2, *The results*, Cambridge University Press, Cambridge, pp. 3-40.
- Leech G. (1998), “Preface: Learner corpora: what they are and what can be done with them”, in Granger S. (ed.), *Learner English on Computer*, Addison Wesley Longman, London-New York, pp. XIV-XX.

- Leto F. (2023), “Costruire e analizzare un corpus di interlingua all’interno di un corso di italiano LS a studenti sinofoni: un modello tra ricerca e didattica”, in *Italiano LinguaDue*, 2, pp. 84-103:
<https://riviste.unimi.it/index.php/promoitals/article/view/21940>.
- Loewen S. (2015), *Introduction to instructed second language acquisition*, Routledge, New York.
- Long M. (1981), “Input, interaction and second language acquisition”, in Wintz H. (ed.), *Native language and foreign language acquisition: Annals of the New York Academy of Sciences*, 379, pp. 259-278.
- Long M. (1996), “The role of the linguistic environment in second language acquisition”, in Ritchie W., Bhatia T. (eds.), *Handbook of second language acquisition*, Academic Press, New York, pp. 413-468.
- Long M. (2020), “Optimal input for language learning: genuine, simplified, elaborated, or modified elaborated?”, in *Language Teaching*, 53, 2, pp. 169-182.
- Mauri C., Ballarè S., Gorla E., Cerruti M., Suriano F. (2019), “KIParla corpus: a new resource for spoken Italian”, in Bernardi R., Navigli R., Semeraro G. (eds.), *Proceedings of the Sixth Italian Conference on Computational Linguistics*, Bari, Italy, November 13-15, 2019: <https://ceur-ws.org/Vol-2481/paper45.pdf>.
- Marcantonio A., Pretto A. (1988), “Il nome”, in Renzi L., Salvi G., Cardinaletti A. (eds.), *Grande grammatica italiana di consultazione*, vol. I, *La frase. I sintagmi nominale e preposizionale*, il Mulino, Bologna, pp. 316-324.
- Milton J., Chowdhury N. (1994), “Tagging the interlanguage of Chinese learners of English”, in Flowerdew L., Tong A. K. K. (eds.), *Entering Text. Language Centre*, The Hong Kong University of Science and Technology, pp. 127-143.
- Nemser W. (1971), “Approximative systems of foreign language learners”, in *International Review of Applied Linguistics*, 9, pp. 115-123.
- Nesselhauf N. (2004), “Learner Corpora and their Potential in Language Teaching”, in Sinclair J. (ed.), *How to Use Corpora in Language Teaching*, John Benjamins, Amsterdam, pp. 125-152.
- Nuzzo E. (2010), “Valutare l’interlingua di studenti cinesi”, in Rastelli S. (ed.), *Italiano di cinesi, italiano per cinesi*, Guerra Edizioni, Perugia, pp. 223-238.
- Nuzzo E., Grassi R. (2016), *Input, output e interazione nell’insegnamento delle lingue*, Bonacci Editore, Torino.
- Paradis M. (2004), *A Neurolinguistic Theory of bilingualism*, John Benjamins, Amsterdam-Philadelphia.
- Paradis M. (2008), *Declarative and Procedural Determinants of Second Languages*, John Benjamins, Amsterdam-Philadelphia.
- Perdue C. (1993), *Adult language acquisition: Cross-linguistic perspectives*, Cambridge University Press, Cambridge, 2 voll.
- Piske T., Young-Scholten M. (2009), *Input matters in SLA*, Multilingual Matters, Bristol.
- Rebuschat P. (2013), “Measuring implicit and explicit knowledge in second language research”, in *Language Learning*, 63, pp. 595-626
- Renzi L., Salvi G., Cardinaletti A. (eds.) (1988), *Grande grammatica italiana di consultazione*, vol. I, *La frase, i sintagmi nominale e preposizionale*, il Mulino, Bologna.
- Sacks H., Schegloff E., Jefferson G. (1974), “A simplest systematics for the organization of turn-taking for conversation”, in *Language*, 50, pp. 696-735.
- Salvi G., Vanelli L. (2004), *Nuova grammatica italiana*, il Mulino, Bologna.
- Selinker L. (1972), “Interlanguage”, in *IRAL (International Review of Applied Linguistics in Language Teaching)*, 10, 3, pp. 209-231.
- Serianni L. (1988), *Grammatica italiana. Italiano comune e lingua letteraria*, in cooperation with Castelvocchi A., UTET, Torino.

- Schmidt R. (2000), “Attention”, in Robinson P. (ed.), *Cognition and second language instruction*, Cambridge University Press, Cambridge, pp. 3-32.
- Spina S. et al. (2006), *Corpus di Italiano L2* (Perugia Foreign University): <https://apps.unistrapg.it/cqpweb/>.
- Sū X. C. (2000), “同形词与‘词’的意义范围——析《现代汉语词典》的同形词词目 *Tóngxíng cí yǔ “cí” de yìyì fānwéi——xī “xiàndài hànyǔ cídiǎn” de tóngxíng cí cí mù, císhū yánjiū*. (Homographs and the meaning of “cí”. Analysis of the homograph entries in *Modern Chinese Dictionary*)”, in *辞书研究 (Dictionary Studies)*, 5, pp. 29-38.
- Tono Y. (2003), “Learner Corpora: design, development and application”, in Archer D. Rayson P., Wilson A., McEnery T. (eds.), *Proceedings of the Corpus Linguistics 2003 Conference (CL 2003), Technical Papers*, 16, University Centre for Computer Corpus Research on Language, Lancaster, pp. 800-809.
- Ullman M. (2001), “The neural basis of lexicon and grammar in first and second languages: The declarative/procedural Model”, in *Bilingualism: Language and Cognition*, 4, 2, pp. 105-122.
- Ullman M. (2005), “A Cognitive Neuroscience perspective on Second Language Acquisition: The declarative/procedural Model”, in Sanza C. (ed.), *Mind and Context in Second Language Acquisition*, Georgetown University Press, Washington, pp. 141-178.
- Valentini A. (1990), “Genere e numero in italiano L2”, in Berretta M., Molinelli P., Valentini A. (eds.), *Parallela 4: Morfologia/Morphologie*, Gunter Narr Verlag, Tübingen, pp. 335-345.
- VanPatten B. (1990), “Attending to form and content in the input: An experiment in consciousness”, in *Studies in second language acquisition*, 12, 3, pp. 287-301.
- VanPatten B., Williams J. (eds.) (2015²), *Theories in second language acquisition: An introduction*, Routledge, New York.
- VanPatten B. (2017), “Situating instructed language acquisition: Facts about second language acquisition”, in *Instructed Second Language Acquisition*, 1, 1, pp. 45-59.
- White L. (1987), “Against comprehensible input: The input hypothesis and the development of second language competence”, in *Applied Linguistics*, 8, pp. 95-110.

WEB REFERENCES

Speechnotes: <https://speechnotes.co/#>.

Sketch Engine: <https://www.sketchengine.eu/>.

Università per Stranieri di Perugia, CELI: <https://www.unistrapg.it/it/certificati-di-conoscenza-della-lingua-italiana/celi-certificati-di-lingua-italiana>.

Università per Stranieri di Siena Centro Cils: <https://cils.unistrasi.it/home.asp>.

Università per Stranieri di Perugia (website corpora): <https://apps.unistrapg.it/cqpweb/>.

Corpus KIParla. L'italiano parlato e chi parla italiano: <http://kiparla.it/>.

APPENDIXES

Appendix 1

Information of 23 prepared recordings based on structured tasks – date, topic, typology and hours of lessons attended.

Date	Topic	Stimulus questions	Hours of lessons attended
Mar.5	1. The weather	What do you see in the picture? What's the weather like? What are these people doing?	204
Mar.19	2. Age and work	Who are these people? How old are they? Where are they? What do they do? What is their job like?	232
Mar.26	3. On the phone	Who are these people? What are they like? What mood are they in? What day is it? Why are they on the phone?	246
Apr.9	4. Family	Who are these people? Why are they taking this picture? Describe them.	274
Apr.16	5. Looking for a house	What is the house in the picture like? Who lives there? How much is the rent? Why do they want to change house? What do they want their new home to be like?	288
Apr.23	6. Bar and restaurants	Who are these people? Why are they eating in a restaurant? What is the restaurant like? Where is it located? What do they order?	302
May 7	7. Supermarket and shops	Tell the story using the past tense. (with 8 images extracted from a video)	330
May 14	8. At the police station	Who is the person in the picture? Where does he work and live? What is he looking at now? How does he feel? Who is he going to ask for help? What is he going to do next?	344
May 28	9. The post office and the bank	Who are the people in the pictures? Are they going to the post office or to the bank? Why? What are they going to do?	372
Jun. 4	10. At the hospital/ pharmacy	What ailment is affecting these people? Where are they going? What is the doctor's advice?	386
Jun.11	11. Beauty care	Today is a big day for 2 of the people in the pictures. Tell what they have done before to give an almost perfect impression.	400
Sep. 10	12. Passport and visa	Your friend is going to Italy to study at the university. Tell him/her what he/she needs to do before leaving. For information on the	414

		check list visit the Italian embassy website (link given).	
Sep. 17	13. Station and airport	Do you prefer to travel by train or by plane? Why? What are the advantages? Motivate your choices.	428
Sep. 24	14. City transport	Tell something about public transport in your city. How do people get around in your city? Where can you buy tickets? Are there problems with traffic? When, where, why?	442
Oct. 8	15. Asking for directions	When you are in an unknown city, do you prefer asking people for directions or do you make use of technological devices? Explain your choice and give examples	470
Oct.22	16. Rent a car	Is it popular in China to rent a car? Who rents a car? In which situations? What are the advantages and the disadvantages?	498
Oct.29	17. Hotel	When choosing a hotel, what do you care more about? Motivate your choice with examples (true or invented)	512
Nov.5	18. Studying	How often do you go to the library? What services does the library in your university offer? Explain to an Italian friend what he/she needs to do in order to use these services and the rules in the library.	526
Nov.12	19. Sport	How often do you practise sports? Where do you usually practise sports? In your opinion, how can you make people understand the importance of sport?	540
Nov.26	20. Amusements/ having fun	How do young people have fun? When, where and with whom do they have fun? What did your parents do to have fun? What has changed?	568
Dec.3.	21. Festivities and holidays	What are the most important traditional festivities in China? In which holidays do people travel? In China do people often go to the seaside or the mountains? If not, what are the preferred destinations for young people? Tell and give examples.	582
Dec.10	22. Appointments/ invitations and dates	In your city what are the wedding traditions?	596
Dec.17	23. Looking for a job	What are the preferences of young people when looking for a job? And what are the difficulties? And what are your preferences?	610

Appendix 2: Information of 23 prepared recordings based on structured tasks – student, duration, number of tokens

21 students named with the letters A to U; the duration of each recording calculated in seconds; “/” indicates two different recordings with the same topic of one student; “-” indicates a missing recording.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1 time	55	42	50	65	80	60	-	55	61	63	57	44	51	56	50	61	60/63	60	-	65	64
token	66	72	81	90	146	102	-	90	105	128	104	57	70	86	70	99	81/82	102	-	111	103
2 time	52	44	60	62	64	64	38	73	47	56	104	57	51	62	60	61	63	60	53	73	63
token	83	95	115	87	133	117	59	104	71	103	96	87	72	105	92	98	98	89	81	110	104
3 time	54	51	62	62	113	71	51	65	58	71	104	62	54	56	48	57	69	61	56	72	26
token	89	101	123	87	236	139	89	115	112	130	106	94	91	107	73	108	105	99	100	108	104
4 time	-	63	152	95	217	89	74	94	69	73	157	103/51	86	60	77	65	105	82	74	76	80
token	-	131	304	150	422	170	154	184	154	141	168	151/84	140	132	123	115	204	164	127	129	146
5 time	56	47	143	95	238	97	77	112	65	60	137	118	55	78	66	63	110	83	87	73	70

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
token	79	96	272	150	503	178	151	193	125	132	138	165	91	174	112	102	179	152	157	119	126
6 time	53	54	99	62	253	83	68	98	65	78	151	153	66	100	77	63	62	81	35	84	65
token	68	109	179	105	513	159	127	169	118	144	149	198	88	183	107	110	103	147	52	147	108
7 time	77	49	96	76	119	62	75	127	62	63	142	83	91	62	-	72	61	77	64	86	94
token	127	97	178	133	281	123	138	234	118	152	157	116	139	140	-	146	124	146	119	144	201
8 time	52	46	90	88	105	62	69	153	73	70	119	85	87	80	71	65	70	67	54	75	67
token	72	101	181	154	236	130	127	279	143	163	120	117	130	160	127	120	125	122	93	123	131
9 time	43	50	85	81	136	69	63	108	75	61	-	93	80	46	64	61	65	63	60	78	74
token	77	101	169	137	317	132	123	209	156	142	-	135	122	105	110	114	139	106	93	132	157
10 time	57	50	76	76	110	70	64	79	59	71	127	75	56	61	57	60	60	65	-	75	69
token	107	103	156	148	244	125	132	165	116	160	144	127	96	153	110	127	129	122	-	126	148
11 time	51	46	70	62	83	87	72	68	67	60	132	87	71	67	61	64	67	76	-	79	79
token	88	95	139	118	201	148	135	133	134	135	149	150	116	159	113	128	135	127	-	133	164

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
token	88	95	139	118	201	148	135	133	134	135	149	150	116	159	113	128	135	127	-	133	164
12 time	61	52	162	64	134	93	74	92	82	61	155	86	68	89	87	83	72	83	59	109	69
token	100	113	315	109	253	163	130	173	158	109	160	133	121	154	135	141	127	147	93	165	143
13 time	64	54	77	62	148	67	62	77	74	58	131	63	73	40	61	67	72	-	56	74	62
token	112	113	161	104	323	128	135	162	153	114	148	101	118	68	109	134	134	-	100	141	140
14 time	73	60	116	65	140	85	76	85	72	63	138	86	103	74	78	71	99	110	89	76	73
token	117	136	250	121	329	181	129	178	161	127	155	160	170	167	146	156	175	192	148	136	168
15 time	86	62	120	76	156	90	67	89	102	63	165	120	79	91	85	84	129	113	93	113	78
token	136	122	209	156	310	191	117	170	214	123	191	207	122	210	148	170	232	189	153	199	180
16 time	74	53	133	63	146	89	62	107	69	81	125	77	104	60	70	73	87	84	72	90	75
token	115	108	255	118	315	181	111	208	133	148	153	136	155	127	118	141	173	145	134	147	151
17 time	90	58	95	71	139	87	72	82	78	75	155	75	67	62	63	132	77	94	113	82	66
token	154	123	190	141	318	193	131	163	176	148	190	118	116	145	109	135	170	168	168	151	145
18 time	83	52	143	73	247	121	71	98	109	85	215	92	74	67	69	89	94	113	100	77	82

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
token	140	115	283	149	443	241	136	164	223	162	235	171	133	124	129	163	213	218	154	164	179
19 time	60	54	140	60	169	92	63	94	87	77	136	101	76	60	63	76	79	80	91	79	79
token	108	117	239	119	347	188	118	160	159	157	155	174	123	127	115	155	149	145	160	151	615
token	108	117	239	119	347	188	118	160	159	157	155	174	123	127	115	155	149	145	160	151	165
20 time	68	-	122	60	237	105	68	99	106	68	78	97	99	59	65	68	103	90	71	81	71
token	123	-	251	103	425	214	137	187	193	147	183	162	154	133	122	141	238	163	129	170	151
21 time	73	52	100	64	179	111	64	101	97	67	69	115	80	73	70	70	86	90	82	84	61
token	138	116	196	137	335	21	122	186	180	151	154	180	147	156	135	150	160	166	147	178	144
22 time	81	53	122	65	163	86	63	77	86	67	66	84	99	67	69	62	98	75	77	97	76
token	128	108	211	121	339	169	120	146	180	140	147	148	166	141	122	130	171	133	126	195	155
23 time	80	56	134	61	221	102	63	84	74	68	165	95	78	66	70	66	80	84	63	93	68
token	132	127	236	121	360	167	113	152	158	128	184	175	151	131	123	132	151	161	110	194	148

Appendix 3

Information of 2 oral exam recordings: date, name, topic, duration, number of tokens

	Exam 1: July 4 th , 2021			Exam 2: July 6 th , 2022		
student	topic	time	token	topic	time	token
A	Beauty care	4m42s	374	Will translators and interpreters have a brilliant future in their careers?	4m07s	154
B	Age and work	6m03s	318	Many people work overtime every day and are always stressed. What are the main causes?	3m49s	150
C	Beauty care	6m11s	433	Many people work overtime every day and are always stressed. What are the main causes?	5m50s	304
D	Supermarket and shops	4m57s	228	Do you have a strong relationship with Chinese traditional festivals?	3m20s	181
E	The post office and the bank	6m45s	505	Do you like online courses? Why?	6m19s	337

student	topic	time	token	topic	time	token
F	The weather	4m52s	349	Why do so many university students want to continue their study in Master courses although they do not have the passion for doing research?	6m59s	414
G	At the police station	5m31s	391	Is it important to do some work part-time while studying at university?	2m08s	133
H	Family	8m33s	393	Do you have brothers or sisters? What is your opinion on the number of children in a family?	3m04s	173
I	The weather	5m23s	251	Do you like online courses? Why?	3m40s	248
J	Looking for a house	6m10s	261	Do you have a strong relationship with Chinese traditional festivals?	3m34s	120
K	Bar and restaurants	6m40s	526	What is your opinion on marriage? For love or for money?	5m32s	337

student	topic	time	token	topic	time	token
L	The weather	7m32s	304	What is your opinion on marriage? For love or for money?	5m46s	229
M	Family	7m50s	365	How does the pandemic influence your life?	5m43s	318
N	At the police station	6m10s	380	Do you have brothers or sisters? What is your opinion on the number of children in a family?	5m50s	334
O	Age and work	5m22s	198	Will translators and interpreters have a brilliant future in their careers?	2m58s	87
P	The post office and the bank	6m46s	330	What is your opinion on the fashion of cosmetic surgery among university students?	6m07s	268
Q	Family	5m13s	359	Is it important to do some work part-time while studying at university?	3m11s	222
R	On the phone	4m39s	413	Do you have a strong relationship with Chinese traditional festivals?	6m18s	368

student	topic	time	token	topic	time	token
S	Bar and restaurants	6m37s	434	Many people work overtime every day and are always stressed. What are the main causes?	3m09s	195
T	At the hospital/ pharmacy	8m07s	416	Is it important to do some work part-time while studying at university?	5m51s	253
U	On the phone	7m26s	330	What is your opinion on the fashion of cosmetic surgery among university students?	7m04s	301

