Connecting Chinese Written Discourse: A Case Study Based on Italian Learners

Chiara Romagnoli
Università degli Studi Roma Tre
chiara.romagnoli@uniroma3.it

The definitions of discourse markers (DMs) provided so far have underlined their high frequency in spoken language, the internal variety of these items and their usage as linking devices between different discourse segments and conversation turns. This contribution is focused on the last feature, i.e. the usage of markers as connectives, and takes into account the written productions of 32 Italian learners of Chinese. A corpus of 78 texts has been created in order to provide a quantitative and qualitative account of the usage of connective devices. Results show how conjunctions and adverbs expressing addition, condition and opposition are preferred over other categories. The statistical analysis carried out also reveals how text types do not affect the usage of connectives. Moreover, no significant difference across time has been noticed, suggesting that much has to be done in order to improve learners' mastery of connective devices in L2 Chinese.

Keywords: Connective devices, Learners Corpus, text type, L2 Chinese

1. *Introduction*

Described as "linguistic, paralinguistic, or nonverbal elements that signal relations between units of talk by virtue of their syntactic and semantic properties and by virtue of their sequential relations as initial or terminal brackets demarcating discourse units" (Schiffrin 1987, 40), discourse markers (DMs) have been labeled and defined in different ways and range from interjections to full prepositional or noun phrases (Celle, Ruarth 2007). The common feature shared by the different definitions of DMs provided in the literature is the linking function ascribed to these items.

Chinese DMs are no exception: although we rarely find them mentioned in pedagogical materials and research on Chinese DMs cannot yet rely on a high number of studies, these items display linking function, are extensively used in spoken language, and include a wide variety of forms. The last features have certainly not helped in classifying Chinese DMs, nor in providing an all-encompassing approach for their description. Among the most recent contributions, Xu (2015, 96), defines Chinese DMs as "lexical items, be they words or short phrases, which help connect prior and following discourse segments". The same source lists four functions of DMs: one of them is connecting two lexico-grammatical constituents and is the one the present study is focused on. In particular, this work in-

vestigates the usage of connectives in the written productions made by Italian learners of Chinese and adopts the definition of *connective* as "each of the invariable forms (conjunctions, phrases, etc.), which indicate relationships that "logically" structure the meanings of the sentence and text" (Ferrari 2010, 271). The Chinese items taken into account mainly include conjunctions and adverbs, but other units of different lexical class and length have also been considered.

This paper is organized as follows: Section 2 briefly reports the classification of connectives adopted in this study, describes a few studies on Chinese connectives and mentions the main features of these items in Italian; section 3, instead, illustrates the main studies carried out in the field of the acquisition of Chinese connectives. In section 4, the research questions, data and methodology of this study are presented whereas the results are reported in section 5, followed by the discussion and pedagogical implications (sections 6 and 7).

2. *Connectives: Classifications and Types*

In recent years, research on connectives and their usage by L2 learners have been frequently carried out from the perspective of discourse grammar and have been based on the central notion of *cohesion* (Halliday, Hasan 1976). The link among units of discourse can be marked explicitly, through conjunctions, adverbs and lexical bundles and implicitly, with lexical or semantic overlap and co-referentiality. This study is focused only on the explicit markers, which are also the only ones presented in Chinese as a foreign language didactic materials and teaching practice.

As for connectives, when understood as DMs, different classifications have been provided. Ferrari (2005) is based on the sematic features of the linking items and is the one adopted in this study since the categories included are all comprised in other works, as those mentioned below. It distinguishes nine categories:

i.   temporal relations;
ii. causal relations;
iii. consequence relations;
iv. concession relations;
v.   condition relations;
vi. linguistic rephrasing;
vii. opposition relations;
viii.addition relations;
ix. *dispositio* relations.

The last one indicates the collocation of sentences within the text and has been labelled as *sequence* in the tables presented in Sections 2.1 and 5.

Focused on the description of discourse, Martin and Rose (2003) propose to distinguish internal and external conjunctions: the former link logical steps internal to the text itself, the latter are instead used to link events in an activity sequence. Most importantly, both types include the same four categories of logical relations: addition, comparison, time and consequence. A further distinction is made on the basis of the dependency status of

the clauses: according to this criterion we may have paratactic, hypotactic and cohesive conjunctions. Although connectives are not limited to conjunctions, this lexical class is the one primarily involved in fulfilling the function of linking segments of discourse.

2.1 Chinese Connectives: Conjunctions and Adverbs

Considering the case of Chinese, conjunctions have been described according to the linked elements, from word to sentence, and according to the relation of dependency between clauses. In Liu et al. (2001), for instance, the distinction is between coordinative conjunctions (*bìngliè liáncí* 并列连词) and subordinative conjunctions (*piānzhèng liáncí* 偏正连词). The same source observes the distinction to be made between conjunction and adverb, claiming that in some cases the role played by adverbs in linking elements is even more crucial than that of conjunctions. The example in (1) illustrates this.

(1)   无论谁听到这个消息都会很高兴。
       Wúlùn *shéi tīngdào zhè gè xiāoxi* dōu *huì hěn gāoxìng.*
       Independently from who listens to this news anybody would be happy.

If the first item, the conjunction *wúlùn* 无论 (independently from), is missing, the meaning does not change; it does if the second element, the adverb *dōu* 都 (all), is not used. Should we then classify adverbs as conjunctions? The answer, according to Liu et al. (2001), is no, but the argument provided is based on a pure syntactic distinction between the two lexical classes, according to which conjunctions can be placed before or after the subject, adverbs only after. However, we know that this is not always the case and the literature has not overlooked the linking function of adverbs nor the variety they display in the position within the clause.

Already Li and Thompson (1981), for instance, in describing the two main categories of linking in Chinese, forward and backward, illustrate the movable and non-movable adverbs having forward and backward-linking function.

More strictly focused on adverbs, Zhang (2014, 319) includes a chapter on those items used to provide cohesion in Chinese, claiming that "in modern Chinese, adverbs play an extremely important, unique role in the organization of texts, especially those disyllabic adverbs usually placed at the beginning of the sentence". The classification proposed by Zhang distinguishes six categories of adverbs and phrases used as connectives. These items have been divided according to the functions they play in expressing i) sequence (*shùnxù* 顺序), ii) inference (*tuīlùn* 推论), iii) explanation (*jiěshuō* 解说), iv) addition (*zhuījiā* 追加), v) shift (*zhuǎnzhé* 转折), and vi) condition (*tiáojiàn* 条件). As we can see from the table 1, in Zhang's articulated classification we can find many adverbs usually labeled as time, degree, negation or modal adverbs (see for instance Liu et al. 2001). Their inclusion in this classification is explained by the author mentioning the role played among clauses, not within the clause, i.e. considering the dimension of the text (*piānzhāng* 篇章).

Table 1 - *Classification of adverbs according to the linking function (modified from Zhang 2014: 311)*

Sequence	Anteriority	先，原先，原来，原本，本来
	Posteriority (sequence)	才，随即，随后，接着，继而，既而，转而
	Posteriority (immediate sequence)	立即，旋即，当即，立刻，顿时，霎时，俄顷，马上
	Beginning	先，处，起先，起初
	End	终于，终归，总归，终究
Inference	Summary	显然，当然，自然，显而易见
	Interpretation	难怪，无怪，怪不得，无怪乎，无怪于，怨不得
	Estimation	也许，或许，兴许，大概，大约
Explanation	Negation	不，其实，事实上，实际上
	Confirmation	真的，的确，确实，诚然
	Integration	原来，本来，果然，果真，果不其然
Addition	Coexistence	也，再，又，还，同时
	Hierarchy	又，再，更，甚至，再者，更有甚者
	Extreme	最，尤，尤其，特别，尤其是，特别是
	Exception	只，就，仅，单，独，光，唯独，独独，偏偏，仅仅
Shift	Opposition	却，仅，倒，反而，反倒，倒是
	Specification	当然，自然，诚然，其实，只是
	Impossibility	无奈，无奈乎，只好，只得
	Unexpectedness	忽然，猛然，蓦然，忽而，忽地，蓦地，兀的，倏地
Condition	Favorable	幸亏，幸好，幸而，幸得，亏得，多亏
	Minimal	至少，至多，起码，不管怎么样
	Unconditional	反正，横竖，横直，左右，高低

Some of the adverbs included in Zhang's classification are extremely frequent and are usually introduced from the very beginning of learning Chinese. Nevertheless, contrary to the attention paid to conjunctions, the linking function of adverbs is rarely highlighted in Chinese didactic material, nor is mentioned the internal variety of adverbs which causes differences both in usage and in the position within the sentence[1]. Another shortcoming in explaining connectives is to limit the description to the sentence level which, although useful, does not reflect authentic usage of language, where "sentences occur in larger contexts, as part of dialogues, monologues, and conversations (as well as written paragraphs)" (Li, Thompson 1981, 631).

2.2 A Brief Comparison Between Italian and Chinese Connectives

Although this study does not investigate Italian connectives as the topic would require much more space, a few lines to describe the linking devices in Italian sentence and text are

---

[1] See for instance *Discover China* (Ding et al. 2010), the handbook used by the target learners of this research.

in order. First of all, the Italian sentence can be understood as a structure including i) clauses linked at the same level or ii) clauses hierarchically linked. In the first case, coordinative conjunctions are used as in (2), in the second one subordinative conjunctions are instead used as in (3)[2].

    (2)   Ho fatto diversi tentativi *ma* non ci sono riuscito.
           I made several attempts but I failed.

    (3)   Verrei *se* potessi.
           I would come if I could.

As a result of the blurred boundary between sentence and text, the list of coordinating conjunctions often includes items used as textual connectives playing different roles and belonging to categories 3 (consequence, as *dunque,* therefore), 6 (rephrasing, as *cioè,* namely), 7 (opposition as *ma,* but), 8 (addition, as *e,* and) of Ferrari's classification mentioned above. As in Chinese, Italian also displays connectives which are mostly used in combination: a typical case is the hypothetical sentence where a conjunction (*se*, if) and an adverb (*allora,* then) are used (De Santis 2019).

    (4)   Se la cosa sta proprio come tu dici, allora va bene.
           If it's just like you say, then fine.

Differently from Chinese, in this type of sentence the Italian connective expressing the condition (*se*) cannot be deleted, whereas the corresponding Chinese conjunctions *rúguǒ* 如果 and *yàoshi* 要是 are often omitted especially in spoken language.

    (5)   你明天不来，我们这个节目就不能表演了。
           Nǐ míngtiān bù lái, wǒmen zhè gè jiémù jiù bù néng biǎoyǎn le.
           Se domani non verrai, il nostro numero non potrà andare in scena.
           If you do not come tomorrow, we cannot perform our show[3].

3. *The Acquisition of Chinese Connectives: Main Studies and Research Results*

In the field of Chinese as a second or foreign language, a number of studies have examined written skills in relation with different proficiency levels and from the perspective of error analysis (see Xiang, Ji 2017). Among the findings within this trend of research, one has revealed that "the occurrence of errors was primarily due to difficulties in establishing coherence in Chinese discourse" (Xiang, Ji 2017, 205). The first studies focused on the usage of connectives and based on the theoretical framework of discourse grammar have underlined the preference for causal connectives by L2 Chinese learners. Cao (2013)

---

[2] Examples 2, 3 and 4 are drawn from the online version of the Italian encyclopedia Treccani available at www.treccani.it (bold in original), last accessed October 20, 2022.
[3] The example is drawn from Abbiati 1998, 260 (Italics in the Italian translation is mine).

and Yang (2013), for instance, explain this finding mentioning different factors related not only with the specific linguistic patterns displayed by the connectives, but also with specific features such as the participants' linguistic background, L1 transfer and task type.

A first attempt to extend the analysis to a large number of connectives is represented by Lu (2019), who sheds light on the distribution of connectives in L2 Chinese and focuses in particular on the difference between those items occurring in pairs and those more flexible which do not necessitate the appearance of another connective.

Lu (2019) provides a valuable contribution to the study of connectives in L2 Chinese for a number of reasons. First of all, it is based on a high number of participants, comprising native speakers, heritage language learners and foreign language learners. Secondly, and differently from the cloze tests widely used to elicit data, the task proposed in this study is a mini-discourse completion task, which seems more adequate than other instruments to provide a context[4]. A third point to be mentioned is the link between learners' background and attention to the discourse-textual layer and the usage of Chinese connectives.

Among the several interesting findings of this study, it is worth mentioning i) the underuse of all connectives considered, ii) the preference for lexical items over connectives to provide coherence, iii) the L1 transfer which leads learners to erroneously perceive the mandatory feature of specific cohesive items, and iv) the more frequent usage of cognitively complex connectives than simple ones by L2 learners. The first result has been connected to learners' (lack of) competence and awareness of the discourse level of language, an aspect which needs to be addressed in order to improve cohesion in L2 Chinese.

Since the notions of cohesion and coherence imply not only the writer but also the reader, it has been also investigated the role played by meta-discourse devices to guide the reader's comprehension.

A step forward in this direction is represented by Liao (2020), which is also focused on L2 Chinese descriptive writing skill and takes into account different proficiency levels. This study describes the textual organizational features at local and global level, i.e. the usage of connectives within and between sentences and clauses and between paragraphs and larger chunks of text. Interestingly enough, it also takes into account the usage of interactional devices such as personal pronouns. The results show correlations between accuracy in the usage of connectives and proficiency level and between the organizational features of the texts written by more advanced participants and the linguistic measures adopted by the scholar to evaluate them. Another interesting finding concerns the lexical complexity, which is positively associated with the accuracy in using meta discourse and interactional devices.

The research studies quoted so far have been based on data collected from learners whose L1 is English. An exception is represented by Romagnoli and Tao (2022), which takes into account the usage of DMs by both Chinese speakers and Italian learners of Chinese. In this account, a broad definition of DMs has been applied and the items analyzed function mainly as connectives at the discourse linking level. Based on oral monologues, this work on the one hand confirms the results illustrated in previous studies, on the other

---

[4] See for instance Bebee, Cummings 2006.

hand also highlights some peculiarities in acquiring Chinese connectives, such as the lack of correlation between proficiency level and usage of the target items and the high number of expressions typically used in written language, despite the oral feature of the task proposed. Although DMs, including the category of connectives, are typically used in spoken language, they can also be found in written language, a dimension which has been mainly investigated from the perspective of error analysis based on data usually drawn from English native speakers. Aim of this study is therefore to fill a research gap presenting a corpus of written data collected from Italian learners of Chinese.

4. *Methodology*

4.1 Research Questions, Data and Hypotheses

The research questions this study aims at answering are:
1. What type of connectives are most frequently used in L2 Chinese productions?
2. Does the usage of connectives change depending on the task type?
3. Does the usage of connectives change across time?

In order to answer these questions, a corpus of written data has been built. The dataset includes 78 texts produced by 32 participants attending the same university course.

All learners are native speakers of Italian, are on average 21 years old and their mastery of Chinese corresponds to the third level of the official Mandarin certification *Hanyu Shuiping Kaoshi*. Their proficiency level has been tested using the end of term written examination, which includes the same tasks and vocabulary of the certification.

The data collected are the written productions uploaded on a voluntary basis by students on the platform Moodle in different periods, from the beginning to the end of the academic year. In particular, students have been asked to freely express themselves on four different topics which can be considered to belong to three types of written tasks: argumentative, descriptive and persuasive.

The number and type of texts uploaded by participants is not the same, therefore for the comparison across time I based my analysis on a selection of texts uploaded by the same participants.

The first three productions have been uploaded during the first term, the last one at the end of the second term. The number of tokens of the first three productions is comparable to that of the last one.

Table 2 - *Dataset*

Text	Text type	Text #	Corpus size
Text 1 What do you think when you listen to the word "China"?	Argumentative	20	1779
Text 2 What is beauty?	Argumentative	18	1879
Text 3 The apps you use	Descriptive	14	1757
Text 4 Give a suggestion to your friend who has just graduated	Persuasive	26	5738

Given the high frequency of markers signaling addition and causal relations in the didactic input, the first hypothesis was to find a higher number of occurrences of these items compared to other connectives. In addition to this, given the co-occurrence of the connectives *se…allora* in Italian hypothetical sentence, I expected learners to mark the Chinese corresponding sentence regularly using the conjunctions *ruguo* or *yaoshi…jiù* 就 (if…then).

Moreover, since one feature of argumentative text is the extensive usage of connectives to signal the paths of an argument, the second hypothesis was to find a higher number of causal, final, consecutive or concessive markers in that type of text. As for the third research question, we expected an improvement due to the linguistic practice from the beginning to the end of the academic year.

4.2 Analysis Procedure

Given the short length of the texts, following Liao (2020) I did not consider semantic and lexical overlap and co-reference. Differently from Liao, I did not take into account intra-clausal cohesion marked by prepositions. My analysis is focused instead on those lexical devices which typically fulfill a linking function in Chinese, that is adverbs, conjunctions, and lexical items linking clauses or adjacent sentences.

In particular, I started processing the texts in order to make them ready for the corpus interrogation. Then I created different corpora and uploaded then on Sketch Engine in order to quantify the data in terms of tokens and types[5]. Using the function "wordlist" I extracted conjunctions and adverbs from the single corpora. Since the automatic extraction provided by the software is not always reliable, all the occurrences have been manually checked and integrated.

Then I compared the corpora according to the research questions this study aims at answering. In particular, I considered the total number of occurrences of connectives in the whole corpus to answer the first research question, I took into account the corpora

---

[5] Available at https://www.sketchengine.eu/, Sketch engine is a software created to collect and analyze large amount of linguistic data, last accessed October 20, 2022.

separately according to text types (three sub-corpora) to answer the second research question and only considered two sub-corpora to compare the difference across time to answer the third research question.

The steps are reported in Table 3.

Table 3 - *Analysis steps*

Step	Procedure	Goal
1) Text processing	Text format conversion, from word to txt	Making texts ready for the corpus interrogation
2) Corpora building	Creation of different corpora and upload on sketch engine	Quantification of the data
3) Corpora interrogation	Extraction of adverbs and conjunctions using wordlist function and quantification	Answer RQ1
4) Corpora comparison based on four corpora	Statistic analysis of the data (Log likelihood test)	Answer RQ2
5) Corpora comparison based on two corpora	Statistic analysis of the data (T- test)	Answer RQ3

In order to quantify the occurrences of connectives in the productions collected, adverbs and conjunctions have been extracted from the corpus. Since the query tool only shows results classified according to different lexical classes, I had to observe each occurrence and further select the target items including only those which have a linking function. The lexical classes provided by the software have been checked against those reported in the *Xiandai Hanyu cidian*: the last one has been the reference to classify the linking devices extracted.

Only adverbs having a linking function and only conjunctions linking phrases or sentences have been included in the calculation.

Moreover, the occurrences of those connectives having more than one lexical class have been distinguished and classified. One such case is *bing* 并 (not at all, and), which as adverb intensifies the negation, as conjunction has instead an additive function as respectively in examples (6) and (7).

(6)   我知道这并不容易。
       Wǒ zhīdào zhè bìng bù róngyì.
       I know that this is not easy at all.

(7)   我想学习很多东西并结识新朋友。
       Wǒ xiǎng xuéxí hěn duō dōngxi bìng jiéshí xīn péngyou.
       I want to learn many things and make new friends.

Whereas (6) has not been considered among the connectives, (7) has been instead included.

The disambiguation has been necessary also for polysemous connectives. A typical case is the conjunction *ér* 而 (and so, but) which can be used to indicate both causal relation and contrast as respectively in (8) and (9).

(8)   中国的菜因为又香又健康而出名。
       Zhōngguó de cài yīnwèi yòu xiāng yòu jiànkāng ér chū míng.
       Chinese food is famous because is good and healthy.

(9)   ·我们花很多时间在手机上，而不是和家人在一起。
       Wǒmen huā hěn duō shíjiān zài shǒujī shàng, ér bù shì hé jiārén zài yīqǐ.
       We spend much time using telephone and not being with relatives.

Not only adverbs and conjunctions display a linking function: verb and phrases such as *bǐfang shuō* 比方说 (for example) and *jiēxiàlái* 接下来 (then) have also been considered as in (10) and (11).

(10)   住在外国有很多的好处，比方说你认识很多人。
       Zhù zài wàiguó yǒu hěn duō de hǎochù, bǐfang shuō nǐ rènshi hěn duō rén.
       Living abroad has a lot of advantages, for examples you (can) meet many people.

(11)   那接下来我来介绍自己的看法吧。
       Nà jiēxiàlái wǒ lái jièshào zìjǐ de kànfǎ ba.
       Then I'm going to introduce my points of view.

5. *Results*

5.1 Frequency of Connectives (RQ1)

First of all, frequency of connectives in relation to the word frequency information in the corpus has been calculated. In table 4, *rank* means the overall ranking of the token, based on frequency of occurrences, among all word tokens found in the corpus, whereas the frequency value indicates its token frequency in the data. Thus, for example, the first item *yīnwèi* 因为 (because) is expressed as 20 and 90, which means that its overall ranking in the corpus is 20 and there are 90 occurrences of *yinwei* in the dataset.

In particular, considering the top 10 markers, we notice how among the first 100 most frequently used words of the corpus, only six are connectives and the raw frequency rate drops after the third one from 63 to 40 occurrences in the whole corpus. On the other hand, the most frequent markers belong to different categories, and this suggests that a variety of semantic and syntactic relations have been marked by different adverbs and conjunctions.

Table 4 - *Connectives frequency in the corpus*

#	*Rank*	*frequency*	*Token*
1	20	90	因为
2	24	76	也
3	30	63	如果
4	48	40	所以
5	69	29	就
6	70	29	但
7	127	17	比如
8	131	16	最后
9	135	16	首先
10	155	13	虽然

In order to verify whether some categories of connectives have been more frequently used than others univariate analysis has been carried out. The choice of this analysis is justified by the inclusion of only one variable (DM category) with 9 levels. The univariate chi-square test assesses the association between observed and expected frequency of the considered variable (Stefanowitsch 2020). The differences observed are always significant, both in terms of tokens (χ^2 = 254.51, df = 8, p < .001) and in terms of types (χ^2 = 13.27, df = 8, p < .01).

In particular, the data reported in Tables 5 and 6 show that additive markers are the most frequently used and also those exhibiting the highest number of different items, the observed frequency surpassing the expected frequency for both tokens and types. The same goes for markers expressing opposition and contrast, which are also frequent but display a more limited internal variety. As for items expressing cause, we find only one conjunction, *yinwei,* showing that learners' vocabulary in this field is very limited.

Table 5 - *Connectives divided by category (tokens)*

	Frequency		
DM category	*Observed*	*Expected*	χ^2
1. Addition	166	66.34	149.72
2. Opposition	93	66.34	10.71
3. Causal	90	66.34	8.44
4. Condition	72	66.34	0.48
5. Consequence	51	66.34	3.55
6. Temporal	47	66.34	5.64
7. Sequence	33	66.34	16.76
8. Rephrasing	29	66.34	21.02
9. Concession	16	66.34	38.20
Total	597		254.51

Table 6 - *Connectives divided by category (types)*

DM category	Frequency		χ^2
	Observed	*Expected*	
1. Addition	12	5.8	6.63
2. Opposition	7	5.8	0.25
3. Causal	1	5.8	3.97
4. Condition	6	5.8	0.01
5. Consequence	5	5.8	0.11
6. Temporal	8	5.8	0.83
7. Sequence	5	5.8	0.11
8. Rephrasing	6	5.8	0.01
9. Concession	3	5.8	1.35
Total	53		13.27

5. 2 Distribution of Connectives (RQ 2)

The picture changes if we instead consider the distribution of connectives in the different text types. As illustrated above, three types have been proposed: argumentative, descriptive and persuasive. In this case the difference in the usage of connectives is not statistically significant and the hypothesis I started from is neither confirmed in terms of tokens (p = 0.13) nor in terms of types (p = .82). In these cases, the comparison was conducted using the loglikelihood test, as the data did not meet the assumptions for the chi-square test (e.g., expected frequencies being smaller than 5).

Moreover, due to the disparity among the sub-corpora in terms of size, the following analysis was conducted on normalised frequencies, calculated on a 1.000-word basis.

Table 7 - *Connectives across different text types (tokens)*[6]

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Text1	14	6	8	1	2	11	3	7	0
Text2	13	11	8	6	3	2	3	3	3
Text3	15	6	8	2	5	5	2	3	1
Text4	15	10	8	10	6	6	4	1	2
Total									

---

[6] Numbers from 1 to 9 in the first line refer to the markers category as reported in tables 5 and 6.

Table 8 - *Connectives across different text types (types)*

	1	2	3	4	5	6	7	8	9
Text1	16	8	2	2	4	10	6	6	0
Text2	10	8	2	4	4	4	6	8	4
Text3	14	8	2	4	4	4	2	4	2
Text4	11	7	1	4	5	9	5	4	3
Total									

5.3 Usage of connectives across time (RQ 3)

Regarding the usage across time, I compared the occurrences of connectives in the first production, delivered during the first term, with those included in the last one at the end of the second term. The reason why I could only select these two productions is that the participants were exactly the same only for these two. This selection caused a drastic reduction of data with only 28 texts available.

The size of the two corpora is very different and the first one (1517 word tokens) is much smaller than the last (3303 word tokens). For this reason, figures have been normalized and although the figures in table 8 may suggest a noticeable improvement from the first to the last text, the difference is neither significant for tokens ($p = .36$) nor for types ($p = .19$).

Table 9 - *Comparison between two texts*

connectives category	Text 1		Text 4	
	Token	Types	Token	Type
Addition	25	8	85	10
Opposition	11	4	55	6
Causal	14	1	46	1
Condition	2	1	55	4
Consequence	4	2	37	5
Temporal	19	5	36	8
Sequence	5	3	21	5
Rephrasing	12	3	7	4
Concession	0	0	10	3

More information about the two corpora is available in Table 10.

Table 10 - *Corpora made of texts 1 and 4*

	Word tokens	*Word types*	*Shortest text*	*Longest text*
Text 1 corpus	1517	418	92	137
Text 4 corpus	3303	633	165	302

6. *Discussion*

The results presented in the previous section lead to some considerations which will be here explained and supported by examples. First of all, learners make wide usage of addition markers, which outnumber the others both in terms of tokens and in terms of types. This result is not surprising given the inclusion of, and the practice on, adverbs such as *yě* 也 (also) and *hái* 还 (in addition) from the very beginning of every Chinese language course and pedagogical material. In addition to these frequent items, we also find words such as *bìngqiě* 并且, *érqiě* 而且, *zàishuō* 再说 (the three all mean: moreover), *yǐjí* 以及 (as well as), *lìngwài* 另外 and *cǐwài* 此外 (besides). Differently from what reported in Romagnoli and Tao (2022) and despite the written nature of the task, learners made no extensive use of markers typical of written language with the exception of *ciwai*, which has however only four occurrences. A point to be noticed is the usage of the comma after or before some addition markers. This is always the case with *ciwai* (12) and often with *erqie* (13) and other markers. This suggests that learners link the usage of these connectives with a pause among segments of discourse.

(12) 此外, 由于空气污染, 城市里经常有很多烟雾 (...)。
      Cǐwài, yóuyú kōngqì wūrǎn, chéngshì lǐ jīngcháng yǒu hěn duō yānwù.
      Moreover, due to air pollution, there is often fog in the city.

(13) 微博对保持与中国名人的联系非常有用, 而且我在微博上交了很
      多朋友 (...)。
      Wēibó duì bǎochí yǔ Zhōngguó míngrén de liánxì fēicháng yǒuyòng, érqiě wǒ zài Wēibó shàng jiāole hěn duō péngyou.
      Weibo is extremely useful to keep in contact with famous Chinese people, in addition I knew many friends on Weibo.

The relation of opposition and contrast has been also often marked but with a more limited inventory of words, which mainly correspond to *dàn* 但 and *dànshì* 但是 (but). The occurrences of *bùguò* 不过 and *kěshì* 可是 (but, however) have been very few with respectively only 2 and 1 occurrences. Surprisingly enough, the conjunction *rán'ér* 然而 with the same meaning, has been more often employed and always followed by a pause as in (14).

(14) 然而, 由于冠状病毒, 你最好留在意大利学习 (...)。
      Rán'ér, yóuyú guānzhuàng bìngdú, nǐ zuì hǎo liú zài Yìdàlì xuéxí.
      However, because of Covid, it would be better if you study in Italy.

Despite the high number of occurrences of the causal connectives, there are only two types: *yinwei* and *yóuyú* 由于, and the first one greatly outnumbers the second.

These results are in line with Yang (2013) who reports an overuse of adversative, additive and causal connectives and partially confirm Liao (2020)'s findings, whose data demonstrate the frequent usage of markers signaling addition, contrast and cause relations, although Liao's data are drawn from participants belonging to different proficiency levels.

More diversified is the inventory of connectives expressing condition: for this semantic relation the most frequently used conjunction is *ruguo* (63), followed by a much lower number of sentences introduced by *yaoshi* (5), *bùguǎn* 不管 (1), *bùrán* 不然 (1), *yīdàn* 一旦 (1), *wulun* (12). In order to verify whether the conjunction has been used in correlation with the adverb *jiu* or the conjunction *nàme* 那么 as in the Italian construction *se...allora*, all the occurrences with *ruguo* have been analyzed but only a few of them, six out of 63, express the hypothetical sentence using the construction.

A few cases of *ruguo* without the meaning of condition have been also reported. In addition to conditional sentence, the Italian conjunction *se* can in fact introduce also an interrogative sentence, so learners add this function to *ruguo* producing agrammatical sentences as that in (15).

(15) *我不知道美的意思是什么, 也不知道如果有一个真的美。
      Wǒ bù zhīdào měi de yìsi shì shénme, yě bù zhīdào rúguǒ yǒu yī gè zhēn de měi.
      I do not know the meaning of beauty, and also do not know if there is a true beauty.

The expression of consequence is mainly fulfilled by the conjunction *suǒyǐ* 所以 (40), rarely by the marker typical of written language *yīncǐ* 因此. Also markers more frequent in spoken language have been used such as *name* in (16).

(16) 如果你看到没有工作机会, 那么, 能继续读硕士。
      Rúguǒ nǐ kàndào méi yǒu gōngzuò jīhuì, nàme, néng jìxù dú shuòshì.
      If you realize that there are not job opportunities, then continue the MA degree.

The relation least frequently marked is the concessive one, with 13 occurrences of *suīrán* 虽然 (although) and only 2 and 1 occurrences of respectively *jǐnguǎn* 尽管 (despite) and *jíshǐ* 即使 (even if). *Suiran* is mostly used in combination with the marker *danshi* as in (17).

(17) 虽然一开始可能会有很多困难, 但是你一定会提高你的语言水平 (...) 。
      Suīrán yī kāishǐ kěnéng huì yǒu hěn duō kùnnan, dànshì nǐ yīdìng huì tígāo nǐ de yǔyán shuǐpíng.
      Although at the beginning there may be some difficulties, then you will increase your linguistic level.

As showed in the previous section, the differences in the usage of markers expressing different semantic relations are significant and this suggests that it is advisable to make the inventory of markers richer, presenting more types belonging in particular to the categories of causal and concessive relations.

Data have also indicated how the usage of markers does not vary, in statistical terms, depending on the text type. A possible explanation of this result is the general lack of explicit instructions about the markers to be used in different text types. We do not find, for instance, similar indications in textbooks and pedagogical material. For written productions, we usually find instructions on the number of characters to be used, on the lexical words related to specific domains or semantic fields, not about functional words as those used to link fragments of discourse.

Nevertheless, it is worth noting that some figures related to the occurrences are very similar (addition or sequence) or even the same (causal relation). Other figures are instead rather different across text types. As previously indicated, in the last text type, that belonging to the persuasive category, learners were requested to give suggestions about the choice to be made after graduation. In this text, the expression of condition is for instance much more frequently marked than in the other ones, as in (18) because the writer often presents different options to the receiver and adopts the conditional form to do it.

(18)   如果你选择去中国, 你会直接接触到中国的文化(. . .)。
        Rúguǒ nǐ xuǎnzé qù Zhōngguó, nǐ huì zhíjiē jiēchùdào Zhōngguó de wénhuà.
        If you choose to go to China, you will directly come in contact with the Chinese culture.

As for the markers expressing temporal relations, a striking difference emerges between the first and the other texts: we find many more different expressions to mark posteriority, anteriority and contemporary events as in (19) in the first argumentative text than in those delivered later.

(19)   上大学以后, 我更了解中国(. . .)。
        Shàng dàxué yǐhòu, wǒ gèng liǎojiě Zhōngguó.
        After having enrolled the university, I understand China more.

Another difference emerges comparing the first two texts to the other ones for the usage of markers expressing rephrasing. The two texts belong to the argumentative type and we find here more types and tokens of this category of markers than in the other two categories, despite the fact that the corpus size of the last text is the biggest one. We can therefore conclude that, although the differences are not statistically significant, the argumentative text type triggers the usage of more markers of rephrasing, also typical of natural conversation as in (20).

(20) 叫美的也不一定是一样的。比如说意大利人喜欢晒太阳，中国人
比较喜欢白皮肤（...）。
Jiào měi de yě bù yīdìng shì yīyàng de. Bǐrú shuō Yìdàlìrén xǐhuan shài tàiyáng,
Zhōngguórén bǐjiào xǐhuan bái pífū.
What is called beauty is not necessarily the same. For instance, Italian like to
tan, Chinese prefer white skin.

The last consideration is related to the comparison between different text types, the first and the last one. As previously mentioned I had to select only a small number of productions written by the same learners at the beginning and at the end of the academic year. The first text type belongs to the category of argumentative text, the fourth to the category of persuasive writing and the corpus made of the last one is considerably bigger than the one composed by the first productions. Nevertheless, in statistical terms there is no significant difference between the two corpora. Using the query tool Voyant I checked if the longest productions were also those containing the highest number of cohesive devices[7]. This correspondence has been confirmed in one case, for the text coded as 1-3, whose number of words is the highest among the first productions selected. Interestingly enough, in one of the shortest texts (1-8) the number of occurrences of connectives was among the highest ones. As for the fourth text type, we notice that in two cases a high number of markers correspond to long texts, i.e. the longest the text, the higher the number of connectives (4-2 and 4-3), whereas one text which is not particularly long has the highest number of markers. This feature therefore supports the lack of statistical difference among the two texts, and the fact that there is not improvement in the mastery of this aspect of vocabulary.

Table 11 - *Comparison between vocabulary and connectives*

Text 1			Text 4		
Production code	Connectives	Length	Production code	Connectives	Length
1-3	8	137	4-2	16	302
1-8	6	98	4-3	18	267
1-12	8	116	4-6	20	228

7. *Conclusions and Pedagogical Implications*

This study aims at providing a picture of the usage of connectives by Italian learners of Chinese. The data, consisting of 78 written productions delivered by students in different times, have been used to i) quantify adverbs, conjunctions and other lexical items with linking functions in Chinese and to ii) classify them into different semantic categories.

Three different text types have been proposed to elicit the data in order to observe whether the text type, argumentative, descriptive or persuasive, could somehow encourage

---

[7] Available at https://voyant-tools.org/, Voyant is a web-based application to perform text analysis, last accessed October 20, 2022.

the usage of specific markers. The written productions have been used to build corpora of different size to obtain wordlists and frequency lists. Statistical analysis of the data has been also carried out in order to verify whether the differences among the usage of markers differed depending on the i) marker category, ii) text type and iii) time.

The results show that the difference in the usage of the markers belonging to the nine categories illustrated above is significant and markers expressing addition are not only the most frequently used, but also those exhibiting the highest number of different items. Markers expressing opposition and contrast are also frequent despite having less internal variety.

As for the correlation between connectives and text type, in this case the difference in the usage of connectives is not statistically significant. This result can be explained by the i) size of the corpora used to elicit data and by ii) the lack of specific instructions related to the written production ability.

Not significant, in statistical terms, is also the difference between the first and last text, although the corpus size of the latter is much bigger than the former. Beside the considerations mentioned to explain the previous result, it has to be noticed that practice on persuasive texts has been less frequently proposed and trained, therefore the inventory of markers turned to be not particularly rich.

In conclusion, this study suggests that more attention should be paid to connectives in Chinese texts, not only providing explicit indications of the notion of cohesion itself, but also improving the metalinguistic knowledge about the role played by adverbs and lexical items other than conjunctions in linking parts of text and discourse. This last task can be carried out through noticing activities and focus on form instructions in order to underline the importance of less frequently presented items and to encourage learners to autonomously produce not only grammatical but also coherent texts in Chinese.

References

Abbiati, Magda. 1998. *Grammatica di cinese moderno*. Venezia: Cafoscarina.
Beebe, Leslie M., Martha C. Cummings. 2006. "Natural Speech Act Data Versus Written Questionnaire Data: How Data Collection Method Affects Speech Act Performance. In *Speech Acts across Cultures*, edited by Gass Susan, Joyce Neu, 65–88. Berlin: Mouton de Gruyter.
Cao, Fei. 2013. "Preference and Constraints by Beginning Learners in the Acquisition of Causal Textual Markers." *Journal of the Chinese Language Teachers Association* 48 (2): 45–67.
Celle, Agnès, Ruth Huart. 2007. *Connectives as Discourse Landmarks*. Amsterdam/Philadelphia: John Benjamins.
De Santis, Cristiana. 2019. "La frase complessa in una prospettiva valenziale." *Italiano LinguaDue*, 2: 285- 300.
Ding, Anqi. 2010. *Discover China*. London: Macmillan.
Ferrari, Angela. 2005. "Le trame logiche dei notiziari accademici." In *Rilievi. Le gerarchie semantico-pragmatiche di alcuni tipi di testo*, edited by Angela Ferrari, 245-290. Firenze: Cesati.
Ferrari, Angela. 2010. "Connettivi." In *Enciclopedia dell'italiano Treccani*, a cura di Raffaele Simone, 271-273. Roma: Istituto dell'Enciclopedia italiana Treccani.
Halliday, Michael A.K., Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.

Li, Charles N., Sandra A. Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California press.

Liao, Jianling. 2020. "Metadiscourse, Cohesion, and Engagement in L2 Written Discourse." *Languages* 5 (25): 1-21.

Liu, Yuehua 刘月华, Pan Wenyu 潘文娱, Gu Wei 故韡. 2001. *Shiyong xiandai Hanyu yufa* 实用现代汉语语法 [Practical grammar of modern Chinese]. Beijing: Commercial press.

Lu, Yuan. 2019. "L2 Distribution of Chinese Connectives: Towards a Comprehensive Understanding of a Discourse Grammar." *Second Language Research* 35 (4): 557-586.

Martin, James Robert, David Rose. 2003. *Working with Discourse: Meaning Beyond the Clause*. Beijing: Beijing University press.

Romagnoli, Chiara, Hongyin Tao. 2022. "Discourse Markers in Mandarin L1 and Italian L2 Monologue Production and Their Pedagogical Implications." In *Pedagogical Grammar and Grammar Pedagogy in Chinese as a Second Language*, edited by Fangyuan Yuan, Baozhang He and Wenze Hu, 167-183. London: Routledge.

Schiffrin, Deborah. 1987. *Discourse Markers*. Cambridge: Cambridge University Press.

Stefanowitsch, Anatol. 2020. *Corpus Linguistics: A Guide to the Methodology*. Berlin: Language Science Press.

Xiang, Catherine, Yue Ji. 2017. "An Analysis of Advanced L2 Chinese Learners' Common Errors and Their Perceptions of Errors in Argument-Based Essays." In *Teaching and Learning Chinese in Higher Education- Theoretical and Practical Issues*, edited by Yang Lu, 201-225. London: Routledge.

Xu, Jiajin. 2015. "Discourse Markers." In *Encyclopedia of Chinese Language and Linguistics,* edited by Rint Sybesma. Brill Reference Online. http://dx.doi.org/10.1163/2210-7363_ecll_COM_00000133.

Yang, Chunsheng. 2013. "Textual Conjunctives and Topic-Fronting Devices in CFL Learners' Written Summaries." *Journal of Chinese Language Teachers Association* 48: 71–89.

Zhang, Yisheng 张宜生. 2014. *Xiandai Hanyu fuci yanjiu* 现代汉语副词研究 [Research on modern Chinese adverbs]. Beijing: Commercial press.

Zhongguo shehui kexueyuan yanjiusuo cidian bianjishi 中国社会科学院研究所词典编辑室 (Chinese Academy of Social Sciences, Institute of Linguistics). 2016. *Xiandai Hanyu cidian* 现代汉语词典 [Dictionary of modern Chinese]. Beijing: Commercial press.

Sitography

https://www.sketchengine.eu/ (last accessed October 20, 2022)
www.treccani.it (last accessed October 20, 2022)
https://voyant-tools.org/ (last accessed October 20, 2022)