

# The use of parallel Corpora for a contrastive (Russian-Italian) description of discourse markers: new instruments compared to traditional lexicography<sup>1</sup>

**Anna Bonola**

Università Cattolica del Sacro Cuore  
Milan  
anna.bonola@unicatt.it

**Valentina Nosedà**

Università Cattolica del Sacro Cuore  
Milan  
valentina.nosedà@unicatt.it

## Abstract

**English.** This paper relates to Corpus Linguistics and in particular to parallel corpus linguistics (Borin, 2002), which promotes the use of parallel corpora for studying languages. After briefly presenting the features of an Italian-Russian parallel corpus designed and compiled by the authors of this paper, and after having clarified the reasons why parallel corpora are such a valid aid, compared to traditional lexicography, especially to investigate linguistic structures characterized by a high pragmatic and language-specific content, such as discourse markers, we propose to test the efficacy of the Italian-Russian parallel corpus by presenting two case studies: the Italian discourse marker *allora* and the Russian particle *ved'*.

**Italiano.** Questo lavoro si inserisce nel filone della *Corpus Linguistics* e in particolare in quello della *parallel corpus linguistics* (Borin, 2002), che promuove l'uso dei corpora paralleli nello studio delle lingue. Dopo aver presentato in breve le caratteristiche di un corpus parallelo italiano-russo progettato e compilato dagli autori del presente contributo, e dopo aver chiarito le ragioni per cui i corpora paralleli, in confronto alla lessicografia tradizionale, sono un ausilio molto più valido per indagare strutture linguistiche ad alto contenuto pragmatico e linguospecifiche, come i segnali discorsivi, ci proponiamo di attestare la validità del corpus parallelo italiano-russo presentando due *case study*: il primo in direzione italiano-russo (il segnale discorsivo *allora*), e il secondo in direzione russo-italiano (la particella *ved'*).

## 1 Parallel corpora<sup>2</sup> and linguistic research

Despite the skepticism of early corpus linguists, who refused to use translated texts to draw conclusions about the functioning of a language<sup>3</sup>, nowadays the scientific community has produced countless works that demonstrate how the use of parallel corpora (PC) can have a greater impact in several areas<sup>4</sup>:

- 1) in linguistic research (contrastive, but not only) PC provide a rather solid empirical basis for comparing two or more languages (Johansson, 2003); moreover, the 'translation method' allows to deepen the semantics and functions of a given linguistic structure (Noël, 2003)<sup>5</sup>;
- 2) in *Translation Studies*, since Baker's work (1993), PC have become a fundamental tool for the study of translated texts, treated as a linguistic variety in its own right, worthy of analysis;
- 3) finally, PC have allowed computational linguistics to make progress in the programming of translation software and, more generally, they have favored the development of NLP (Calzolari and Lenci, 2004).

However, these 3 points must be integrated with a further aspect: PC are in fact very useful for the heuristic phase of a contrastive analysis on polyfunctional linguistic elements that are strongly influenced by the context.

---

<sup>1</sup> This paper is the result of a research in which the authors have equally contributed; however, Valentina Nosedà is the author of sections 1, 1.1, 3.1 and 4, Anna Bonola of sections 2, 3 and 3.2.

<sup>2</sup> A parallel corpus consists of texts in a language A, aligned (usually at the sentence level) with the corresponding translations in a language B. If bidirectional, the corpus will also contain language B originals alongside translations in language A.

<sup>3</sup> The reasons for this skepticism can be traced to the generally recognized existence of the so-called universals of translations, to the influence that the source text often exerts on translators and their final product, and to the freedom with which a translator can interpret the source text while transferring its contents into a target text (Olohan, 2004; von Waldenfels, 2012; Zanettin, 2012).

<sup>4</sup> Another field where parallel corpora have proved to be useful is the teaching of second languages, since bilingual corpora, first of all, allow students to grasp equivalences and differences between L1 and L2, thus acquiring greater awareness of the structures of a studied language (Granger, 2003), and secondly help them learn unknown words (Bernardini, 2004).

<sup>5</sup> Noël (2003) was among the first to promote the use of PC not only for contrastive analysis but also to deepen the semantic investigation of one of the two aligned languages. In Russian studies, many works have been carried out following Noël's example, including (Zaliznjak, 2015; Levontina and Denissova, 2017; Zaliznjak, Denissova and Mikeljan, 2018).

In this paper we will show an example of such further use of PC, applying it to the contrastive study of two DMs.

## 1.1 The Italian-Russian parallel corpus of the Russian National Corpus

In Russian studies, the active use of language corpora fell slightly behind the spread of Corpus Linguistics around the world and it has been directly linked to the creation of the Russian National Corpus (*Nacional'nyj korpus russkogo jazyka*, from now on: NKRJa) in 2004. With its 500 million words, its numerous specialized sub-corpora and a highly sophisticated search engine, NKRJa has quickly become an essential tool for the study of Russian<sup>6</sup>.

In 2005 NKRJa already presented a section dedicated to PC, although for the Russian-Italian pair there was only a small pilot corpus, not very balanced and almost useless for any type of research. A first expanded version – resulting from the collaboration between Catholic University of Milan (Università Cattolica del Sacro Cuore di Milano), the University of Bologna (Università di Bologna) and the Russian Language Institute in Moscow (*Institut Russkogo Jazyka imeni V.V. Vinogradova*) – became available in 2015. Now the Italian-Russian PC (it-ru PC) exceeds 4 million words and has become a sufficiently large tool allowing to conduct scientifically valid and statistically relevant research.

The corpus, compiled according to precise criteria, has the following features<sup>7</sup>: i) is bidirectional: it contains Russian originals translated into Italian and vice versa; ii) it includes several literary works and essays (from 19<sup>th</sup>, 20<sup>th</sup> and 21<sup>st</sup> centuries) as well as some newspaper articles written in the last decade (and this variety distinguishes it from other parallel corpora in NKRJa); iii) like all the other sections of NKRJa, it has three types of annotation: metatextual<sup>8</sup>, morphological and semantic.

## 2. The use of parallel corpora for the analysis of discourse markers

A field in which parallel corpus linguistics seems to have great potential, especially if compared to more traditional research methods, is that of discourse markers (DMs), i.e. multi-functional linguistic elements of various origins (adverbs, verbs, particles, etc.) that can operate at a textual, discursive, interactive, modal, social and contextual level<sup>9</sup>. DMs have come to the attention of researchers especially during the eighties, as a result of a new pragmatic direction in language studies, and since then considerable progress has been made in this area<sup>10</sup>. However, the use of electronic corpora in the description of DMs is still in its initial phase.

The difficulty of producing a fully automatic tool for the analysis of DM is due to the fact that these are procedural and multifunctional elements expressing pragmatic and discursive functions which are clarified only in relation to the context or to the communicative situation, whose automatic annotation is still developing<sup>11</sup>. Moreover, syntactically, DMs are optional (can be removed), relatively mobile in the utterance and come from diverse grammatical classes, on which depends their syntactical integration (Crible, 1917: 106).

Therefore, the discussion on the automatic processing of DMs is currently still focused on the “need for functional paradigmatic studies that include every kind of DMs, possibly in multifunctional approaches for better generalization” in order to “provide a solid basis for comparative or contrastive analysis between languages and frameworks” (Crible, 2017: 100).

Some recent experiments for the identification and annotation of DMs are worth noting, like for example (Bolly et al., 2017), even though the empirical method they present is still matching manual and automatic annotation. For a fully automatic cross-linguistic analysis of DMs, which takes into account not only syntaxis

---

<sup>6</sup> A detailed description of the corpus and its sub-corpora (including all the information about the available annotations) can be found on the corpus website ([www.ruscorpora.ru](http://www.ruscorpora.ru)). See also (Aa.Vv 2005) and (Plungjan 2009).

<sup>7</sup> For a description of the Russian-Italian PC and its design criteria, see (Noseda, 2018).

<sup>8</sup> It provides various pieces of information about a text: author, date, genre, number of words, etc.

<sup>9</sup> This list of functional areas summarizes the results of the debate on the classification of DMs – for a review see (Schiffrin 2001; Frediani and Sansò 2017); for the discussion in Italy see (Bazzanella 2001: 41-42) – although we avoid entering into the discussion on labels, whose boundaries are subject to change and have a graduated character (Molinelli, 2018: 277).

<sup>10</sup> For a review of the features of the DMs highlighted by the research, up to the most recent studies see (Frediani and Sansò, 2017).

<sup>11</sup> As far as Russian is concerned, in the NKRJa only the multimodal sub-corpus (4 million words) is pragmatically annotated: the search engine can be interrogated on the basis of specific contexts (at the doctor's, at the restaurant, etc..) and linguistic acts (complaint, prohibition, apology, etc..).

Among Italian corpora, we can name the AVIP corpus (<http://www.parlaritaliano.it/index.php/it/corpora-di-parlato/673-corpus-avip-api>) and PraTID (<http://www.parlaritaliano.it/index.php/it/progetti/35-pratid-un-sistema-di-annotazione-pragmatica-di-dialoghi-task-oriented>), which are fully or partially annotated at a pragmatic level.

(Cinque 1999) but also semantics and pragmatics, the annotation of PC, according to Crible (2017: 107), should consider the following levels: ideational (the relation between real-world events), rhetorical (the relation between epistemic and speech-act events), sequential (the shaping of discourse segments) and interpersonal (speaker-hearer relationship). Therefore, the large amount of data that can be consulted today through electronic corpora, as far as DMs are concerned, has yet to find a way to be processed employing a targeted annotation. More specifically, concerning the automatic analysis of DMs in Russian, some supracorpora databases (SCDB), resulting from the processing of some bilingual parallel corpora within NKRJa, have recently been developed (Zatsman, Inkova, Kruzchkov and Popkova, 2016). Their aim is to increase the functionality of parallel corpora for goal-oriented cross-linguistic research on various linguistic elements. For the moment, there is one SCDB for French-Russian contrastive analysis of verbs (Zatsman and Buntman, 2015) and one for textual connectors (Inkova, 2018).

The it-ru PC used for our research does not have an annotation that takes into account pragmatic and discursive parameters; moreover, we still do not have Russian-Italian SCDB for such particular linguistics elements as DMs. Therefore, for the moment, we tested the effectiveness of it-ru PC as a tool for linguistic analysis in the heuristic phase, as it provides a significant number of examples in a short time, allowing researchers to clarify and adjust their intuition regarding a given research question (corpus-based approach) or to formulate new hypotheses (corpus-driven approach) (Mikhailov and Cooper, 2016: 15-16). If this is generally helpful, it is even more useful for DMs, i.e. linguistic elements that both in Italian and Russian have been developing textual, discursive, modal and pragmatic functions that make them multifunctional and often language-specific, but frequently still lacking an adequate description (Proietti, 2000: 227) (Benigni and Nuzzo, 2019: 152–154)<sup>12</sup>, especially if we consider current lexicography.

As we will show in this paper, the effectiveness of our PC (in its current form) for the heuristic phase of a contrastive corpus-driven or corpus-based approach – well described in (Crible, 2017) – lies in the fact that:

- 1) it makes the multi-functionality of DMs easily emerge, clarifying it by contrast with another language, as description through linguistic comparison, “rend le dispositif d’analyse plus puissant: elle peut suggérer, d’une part, de nouvelles hypothèses pour les faits constatés; elle peut, d’autre part, inciter à réexaminer des hypothèses existantes” (Lamiroy, 1984: 224);
- 2) if a given DM presents recurrent functional equivalences in the language compared, it is possible to determine if in the L2 there are DMs associated with specific functions as well;
- 3) finally, analyzing quantitative data (even with a relatively small number of examples), we can see the preferential strategies of each language to express certain functions, and in some cases, as illustrated in section 3, it is also possible to make some assumptions about possible structural differences between the two compared languages.

In section 3 we will exemplify the abovementioned points by analyzing Italian *allora* and Russian *ved’*, two of the most frequently used DMs in the respective languages.

### 3. The DMs *allora* and *ved’*

Concerning the pragmatic-textual multi-functionality of DMs (section 2, point 1), both Russian and Italian lexicographic descriptions are particularly poor and often do not distinguish contextual elements from the functional core meaning of the DM under investigation.

For example, as far as *allora* is concerned, DISC 2008, among the several dictionaries that we have consulted, is the only one providing some clear categories about the discursive use of this word, which can be a temporal adverb, a conjunction or an actual DM. According to DISC, *allora*, as DM, refers to shared knowledge in dialogues (*Allora?*) or in exhortative, imperative and interrogative sentences (*e allora sei pronto?*). This brief description, although correct, is rather uncomplete and it uses contextual categories, such as sentence or text type, without specifying how their role interacts with the functionality of the DM.

As for traditional Russian lexicography, the description of DMs is not better: in both traditional (Ušakov, 1935) and recent dictionaries (Kuznecov, 2000; Efremova, 2001; Ožegov and Švedova 2003) the particle-conjunction *ved’*, whose various meanings are summarized in (Morozov, 2014: 259), is defined as follows: 1) conjunction in those sentences that indicate the cause or the motivation of a previous statement; 2) concessive conjunction; 3) it expresses a hypothetical or possible state; 4) particle that underlines or contradicts what has

---

<sup>12</sup> In particular, in the article, dedicated to the use of corpora for teaching DMs, the authors underline how even in this field has emerged so far “a lack of contextualization of pragmatic phenomena and a shortage of natural conversational models, exemplifying the real use of language” (Benigni and Nuzzo, 2019: 154).

been said; 5) it emphasizes adversative conjunctions such as *no* [but]<sup>13</sup>, *a* [but, and], *daže* [even]; 6) in conditional clauses it means *togda* [then], *v takom slučae* [in this case]; 7) it indicates a statement from which a conclusion will be drawn; 8) it gives emotional color to spoken language; 9) in questions and exclamations it means *neuželi ne?*, *razve ne* [really/indeed].

Such a functional heterogeneity, as well as the variety of aspects involved, shows that the core meaning of *ved'* provided by lexicographic descriptions is quite vague and even confused. Moreover, as in the case of *allora*, the problem of distinguishing the function of connector from that of DM remains.

Thanks to our corpus-driven analysis in the it-ru PC, a much more precise and complex description has surfaced.

### 3.1 Allora

Our analysis took into account the first 200 occurrences of *allora* automatically extracted from the corpus (100 in Italian originals and 100 in texts translated from Russian)<sup>14</sup>.

Firstly, we considered Russian DMs corresponding to *allora* both in Russian translations and in Russian original texts; secondly, we examined their different functions. Our goal was, on the one hand, to clarify the multi-functionality of *allora* by contrast with Russian, and on the other to compare our results with the descriptions of this DM provided by traditional lexicography and some linguistic research works. This allowed us to verify if our PC, even in its current form, can be useful to integrate these resources towards a more precise description.

As *Allora* is highly polysemic (it combines temporal, logical and pragmatic values) and multifunctional (it can be an adverb, a connector or a DM), we found out that it does not have full functional equivalents in Russian; in fact, quite frequently (25 occurrences) *allora* does not show any equivalent at all: either it is omitted in the Russian translation or it is inserted in the Italian translation without a corresponding DM in the Russian original; its adverbial and connective values are rather carried out by different and thus highly specialized markers with metatextual/metanarrative, interactive and pragmatic functions (this distinction is provided in Bazzanella, 2001) (see section 2, point 2). More precisely:

- *togda* [then] mostly conveys adverbial and connective meanings;
- *značit* and *vychodit* [so] (connectives) express conclusion by inference or deduction;
- *(i) tut* [and then] often expresses temporal correlation and is combined with the metanarrative function typical of *allora*, that marks the different phases of narration.
- *Tak/itak* [so] add two pragmatic functions to the basic consequential meaning: i) interactional function (beginning or end of the interaction, and turn-taking in a conversation); ii) metanarrative function (restarting the narration or marking the narrative phases);
- *Nu* and *že* [well] never have temporal meaning, but they carry out pragmatic functions, emphasizing the interactional process as well as turn-taking. *Nu* and *že* do not seem to express any consequential component.

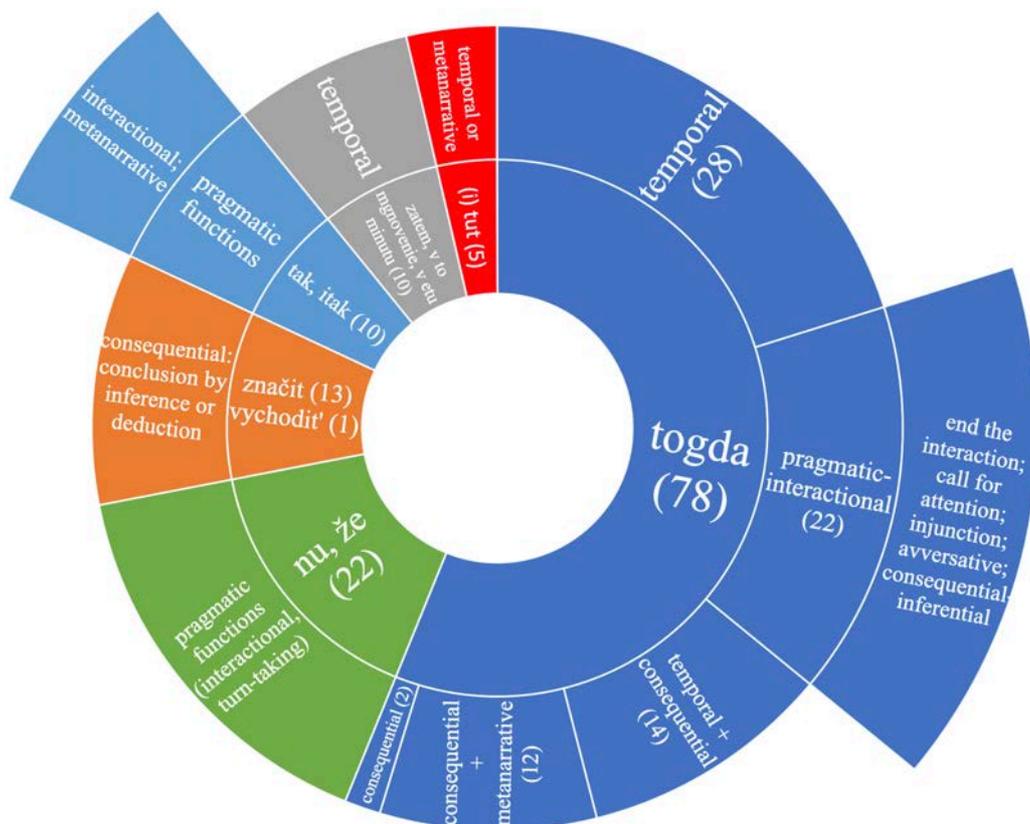
These results are summarized in Figure 1, which shows, in addition, quantitative data. In this regard we must point out that we had to leave out some examples due to translation errors, omissions etc.; as a result, the number of examples that we could actually take into account amounts to 164 (including the 25 cases of zero correspondence which are not showed in Figure 1):

---

<sup>13</sup> The translations that we provide in brackets are approximate since even in English there is never a single equivalent for these words.

<sup>14</sup> This bidirectional approach allows determining if the behavior of a given linguistic unit differs according to text type (i.e. original versions vs translations). In this sense, a bidirectional parallel corpus has proved to be an extremely helpful resource.

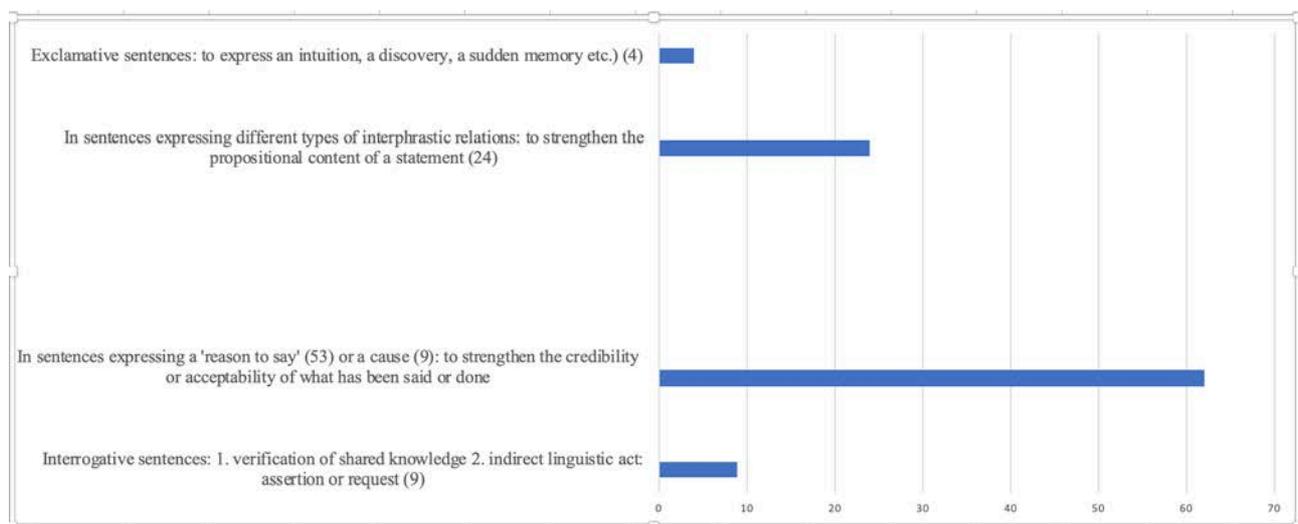
Figure 1: Data resulting from a corpus-driven analysis of DM *allora*



### 3.2 *Ved'*

Biagini and Bonola (2019, in press) have recently applied to *ved'* a similar heuristic method of investigation using the it-ru PC. They examined the first 100 occurrences automatically extracted from the corpus (both in originals and translated texts). In this case, the analysis was carried out considering first of all the contexts of occurrence. The primary goal was to identify the core meaning of *ved'*, in order to distinguish it from other peripheral values. The results of the analysis are summarized in Figure 2, followed by a brief explanation:

Figure 2: Data resulting from a corpus-driven analysis of DM *ved'*



- unlike what is stated in dictionaries, Biagini and Bonola would not attribute to *ved'* the encoding function of clause linking, even though in our corpus the group of contexts which exhibit an interphrastic relation is the second in terms of entity: *ved'* in fact occurs in sentences that express very different kinds of relations (such as adversative and causal), which, nevertheless, in almost all the examples are codified by conjunctions or are inferable from the propositional content of the statements, instead of directly depending on *ved'*.

- secondly, in more than 50% of the analyzed contexts, *ved'* occurs in the presence of two sentences, the second of which expresses a 'reason to say' (i.e. a reason why something was previously said) instead of a mere causal relation. Strengthening the illocutionary force of the second sentence by referring to a shared background that the speaker wants to recall, *ved'* provides the listener with useful hints to overcome the inferential process. In these contexts, *ved'* realizes the macro-functions of expressing textual cohesion (discourse marker), social cohesion and personal attitude (pragmatic marker).

- thirdly, the semantic core of *ved'* (if used when referring to a shared knowledge) persists in particular when it functions as a pragmatic marker that manages social cohesion and modulates illocutionary force (in questions and 'reasons to say') or as an element which favors the inferential process (in 'reasons to say'). When, on the other hand, it carries out the role of discourse marker favoring textual cohesion, it still refers to shared knowledge, but apart from this, nothing else is presupposed.

If the results described above exemplify points 1 and 2 of section 2, concerning the multi-functionality of *allora* and *ved'* or the specialization of their equivalents in Russian (for *allora*) and in Italian (for *ved'*), for point 3 – i.e. the preferential strategies of Russian and Italian regarding the expression of certain discursive functions – it was very useful to analyze the asymmetries emerged from the it-ru PC, i.e. the cases of omission or addition of *allora* and *ved'* in target texts compared to the originals. This analysis showed that both DMs are sometimes omitted in translation or they are added in the absence of a correspondent marker in the original. In addition to this, neither of the two DMs has a perfect functional equivalent in the target language, but they distribute their many functions on partial equivalents. This demonstrates a certain language-specificity of both DMs (on the relationship between the number of translation variants and language-specificity of DMs see Inkova, 2017).

Moreover, as far as *allora* is concerned, we observed that using this marker we tend to give a logical (consequential) interpretation to the temporal relationship between two circumstances: "in that moment/that circumstance" can, in fact, be interpreted through *allora* also as a consequential relationship. Here we can see the preference of Italian for logical cohesion in the text. Russian, on the contrary, often simplifies this temporal-consequential relation in a strictly temporal sense, translating *allora* with temporal adverbs or adverbial phrases (on this difference between Russian and Italian, a consequence of Latin syntax, see Govorucho, 2007).

#### **4 Conclusions: a hypothesis on the differences between Italian and Russian regarding the use of DMs**

Our conclusions regard, firstly, the evaluation of the tool we adopted for our corpus-based contrastive analysis of DMs, i.e. the Russian-Italian bidirectional parallel Corpus of NKRJa. At the moment we can say that this corpus is suitable for the heuristic phase, but it does not yet provide sufficient data to draw general conclusions from a systemic or typological point of view. Any assumption about possible structural differences related to the use of DMs in Russian and Italian should be supported by a larger number of data. Nevertheless, a heuristic analysis allowed us to formulate some preliminary hypotheses.

More precisely, we were able to register the tendency of Russian to express purely pragmatic functions, both cognitive and interactive<sup>15</sup>, through an ancient group of primitive particles, such as *ved'*, *nu*, *že*, which are more specialized if compared to DMs of more recent origin, such as *togda*, which maintains an adverbial and connective function as well. On the contrary, Italian tends to form multifunctional DMs of verbal or adverbial origin which combine their pragmatic features with the task of guaranteeing logical cohesion in the text and interphrastic relations.

This is a broad – and according to us new – observation on a structural difference between the two languages, which deserves to be further explored by investigating – both from a diachronic and a synchronic point of view – more Russian and Italian DMs. All this demonstrates how a heuristic corpus-driven study allows, on the one hand, to quickly obtain linguistic descriptions on the functioning of DMs that are more precise than those provided by traditional tools and, on the other, to open up new hypotheses for wide-ranging research.

---

<sup>15</sup> On this distinction see (Bazzanella, 2001).

## References

- Aa.Vv. 2005. *Nacional'nyj korpus russkogo jazyka: 2003–2005. Rezul'taty i perspektivy*. Indrik, Moskva.
- Mona Baker. 1993. *Corpus linguistics and translation studies – Implications and applications*. Mons Baker, Gill Francis and Elena Tognini-Bonelli (eds). *Text and technology. In honour of John Sinclair*. John Benjamins Publishing, Philadelphia and Amsterdam, pp 233-250.
- Carla Bazzanella. 2001. *Segnali discorsivi e contesto*. Wilma Heinrich and Christine Heiss (eds). *Modalità e Substandard*. CLUEB Bologna, pp. 41–64.
- Carla Bazzanella, Cristina Bosco, Barbara Gili Fivela, Johanna Miecznikowski and Francesca Tini Brunozzi. 2008. *Segnali discorsivi e tipi di interazione*. Cristina Bosisio, Bona Cambiagli, Emanuela Piemontese and Francesca Santulli (eds). *Aspetti linguistici della comunicazione pubblica e istituzionale*. Guerra, Perugia, pp. 239–265.
- Valentina Benigni and Elena Nuzzo. 2018. *L'insegnamento dei segnali funzionali in russo come lingua seconda*. Alberto Manco (ed.) *Le lingue extra-europee e l'italiano: aspetti didattico-acquisizionali e sociolinguistici. Atti del LI Congresso Internazionale di Studi della Società di Linguistica Italiana (Napoli, 28-30 settembre 2017)*. Officinaventuno, Milano, pp. 151-165.
- Silvia Bernardini. 2004. *Corpora in the classroom. An overview and some reflections on future developments*. John Sinclair (ed). *How to use corpora in language teaching*. Benjamins, Amsterdam.
- Francesca Biagini and Anna Bonola. 2019 (in press). *Descrizione semantico-funzionale delle particelle russe e corpora paralleli. Un'analisi contrastiva (italiano-russo) corpus-based di ved'*. Iliyana Krapova, Svetlana Nistratova and Luisa Ruvoletto (eds). *Studi italiani di linguistica slava: nuove prospettive e metodologie di ricerca*. Edizioni Ca' Foscari, Venezia.
- Catherine T. Bolly, Ludivine Crible, Liesbeth Degand and Deniz Uygur-Diestexhe. 2017. *Towards a model for discourse marker annotation. From potential to feature-based discourse markers*. Chiara Frediani and Andrea Sansò (eds). *Pragmatic Markers, Discourse Markers and Modal Particles. New perspectives*. John Benjamins Publishing Company, Amsterdam and Philadelphia, pp. 71-98.
- Lars Borin. 2002. ... *And never the twain shall meet?* Lars Borin (ed.) *Parallel Corpora, parallel worlds. Selected papers from a symposium on parallel and comparable corpora at Uppsala University, Sweden, 22-23 April 1999*. Rodopi, Amsterdam and New York, pp. 1-43.
- Nicoletta Calzolari and Alessandro Lenci. 2004. Linguistica computazionale. Strumenti e risorse per il trattamento automatico della lingua. *Mondo Digitale 2*: 56-69.
- Guglielmo Cinque. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford University Press on Demand.
- Ludivine Crible. 2017. *Towards an operational category of discourse markers: A definition and its model*. Chiara Frediani and Andrea Sansò (eds). *Pragmatic Markers, Discourse Markers and Modal Particles. New perspectives*. John Benjamins Publishing Company, Amsterdam and Philadelphia, pp. 99-124.
- DISC. 1997–2008. *Dizionario italiano Sabatini Coletti*. Giunti, Firenze [http://dizionari.corriere.it/dizionario\\_italiano](http://dizionari.corriere.it/dizionario_italiano)
- Tat'jana F. Efremova, (ed). 2001. *Tolkovyj slovar' služebnyh častej russkogo jazyka*. Russkij jazyk.
- Chiara Frediani and Andrea Sansò (eds). *Pragmatic Markers, Discourse Markers and Modal Particles. New perspectives*. John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Roman Govoruchko. 2007. Složnoe predloženie s vremennym značeniem v ital'janskom i russkom jazykach i problemy rečevogo uzusa. *L'analisi linguistica e letteraria* 15(1): 93–115.
- Sylviane Granger. 2003. *The corpus approach: a common way forward for contrastive Linguistics and Translation Studies?* Sylviane Granger, Jacques Lerot and Stephanie Petch-Tyson (eds). *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Rodopi, Amsterdam and New York, pp. 17–29.

- Ol'ga Ju. In'kova. 2017. Principy opredelenija lingvospecificnosti konnektorov. *Komp'juternaja lingvistika i intelektual'nye tehnologii* 16(2): 150–60.
- Ol'ga Ju. In'kova. (ed.), 2018. *Semantika konnektorov: kontrastivnoe issledovanie*. Torus press, Moskva.
- Sergej A. Kuznecov. (ed). 2000. *Bol'shoj tolkovyj slovar' russkogo jazyka*. Norint, Sankt Peterburg.
- Stig Johansson. 2003. *Contrastive linguistics and corpora*. Sylviane Granger, Jacques Lerot and Stephanie Petch-Tyson (eds). *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Rodopi, Amsterdam and New York, pp. 31-44.
- Béatrice Lamiroy. 1984. *La valeur heuristique de la comparaison linguistique: un exemple concernant le français, l'espagnol et l'italien*. Guillet Alain and Nunzio La Fauci (eds). *Lexique-grammaire des langues romanes. Actes du premier colloque Européen sur la grammaire et le lexique comparés des langues romanes, Palerme, 1981*. John Benjamins Publishing Company, Amsterdam and Philadelphia, pp. 223–230.
- Irina B. Levontina and Galina V. Denissova. 2017. Ital'janskoe magari i ego russkie perevodnye ekvivalenty: raznye diskursivnye strategii. *Komp'juternaja lingvistika i intelektual'nye tehnologii* 16(2): 261–270.
- Mikhail Mikhailov and Robert Cooper (eds). 2016. *Corpus Linguistics for Translation and Contrastive Studies. A guide for research*. Routledge, London and New York.
- Evgenij A. Morozov. 2014. Diskursivnye slova *ved'* i *doch'*: opyt semantičeskogo analiza (na materiale slovarej sovremennogo russkogo i nemeckogo jazykov). *Problemy istorii, filologii, kul'tury* 3: 258–60.
- Dirk Noël. 2003. Translations as evidence for semantics: an illustration. *Linguistics* 41(4): 757–785.
- Valentina Nosedà. 2018. La *corpus revolution* russa e il corpus parallelo italiano-russo: storia, criteri di compilazione e usi. *L'Analisi linguistica e letteraria* 24(2): 115–132.
- Maeve Olohan. 2004. *Introducing corpora in translation studies*. Routledge, London and New York.
- Sergej I. Ožegov and Natal'ja Ju Švedova (eds). 2003. *Tolkovyj slovar' russkogo jazyka*. ITI Technologii, Moskva.
- Vladimir A. Plungjan (ed). 2009. *Nacional'nyj korpus russkogo jazyka: 2006–2008. Novye rezul'taty i perspektivy*. Nestor-Istorija, Sankt-Peterburg.
- Domenico Proietti. 2000. *Comunque dalla frase al testo*. *Studi di grammatica italiana* 19: 175–231.
- Deborah Schiffrin. 2001. *Discourse markers: language, Meaning, and Context*. Deborah Schiffrin et al. (eds), *Handbook of Discourse Analysis*. Blackwell Publishers, Oxford, pp. 54–73.
- Dmitrij N. Ušakov (ed). 1935-40. *Tolkovyj sovar' russkogo jazyka: v 4 t*. Sovetskaja enciklopedija, Moskva.
- Ruprecht von Waldenfels. 2012. Polish tea is Czech coffee: advantages and pitfalls in using a parallel corpus in linguistic research. *Trends in Linguistics* 247: 263–28.
- Anna A. Zaliznjak. 2015. Lingvospecificnye edinicy russkogo jazyka v svete kontrastivnogo korpusnogo analiza. *Komp'juternaja lingvistika i intelektual'nye tehnologii* 13: 683–695.
- Anna A. Zaliznjak, Galina V. Denissova and Irina L. Mikeljan. 2018. Russkoe *kak-nibud'* po dannym parallel'nych korpusov. *Komp'juternaja lingvistika i intelektual'nye tehnologii* 17: 803–817.
- Federico Zanettin. 2012. *Translation-driven corpora*. Routledge, New York.
- Igor' M. Zatsman and Nadezhda Buntman. 2015. Outlining goals for discovering new knowledge and computerised tracing of emerging meanings discovery. Andrea Garlatti and Maurizio Massaro (eds), *16th European Conference on Knowledge Management Proceedings. ECKM 2015*. Academic Publishing International Ltd, Reading, pp. 851–860.
- Igor' M. Zatsman, Ol'ga Ju. In'kova, Michail G. Kružkov and Natalija A. Popkova. 2016. Representation of cross-lingual knowledge about connectors in supracorpora databases. *Informatika i ee Primeneniya (Informatics and its Applications)* 10(1): 106-118.