Macrothink Institute™

**International Journal of Linguistics**
ISSN 1948-5425
2014, Vol. 6, No. 2

# Semantic Classes and Functions of Lexical Presupposition Triggers: An Experimental Investigation of Chatters' Use

Francesca Ferrucci

Dipartimento di Istituzioni Linguistico-Letterarie, Comunicazionali e Storico-Giuridiche dell'Europa (DISTU), Viterbo University (Italy), Via San Carlo, 32 – 01100 Viterbo (VT), Italy

Tel: 39-076-135-7867 E-mail: francesca.ferrucci@informa.pro

Received: March 10, 2014 Accepted: March 19, 2014 Published: April 4, 2014

doi:10.5296/ijl.v6i2.5264 URL: http://dx.doi.org/10.5296/ijl.v6i2.5264

**Abstract**

Prior research suggested the possibility of establishing systematic linkages between some intrinsic features of a presupposition and textual and pragmatic functions that it can carry out with greater probability. This study aims, firstly, to provide an organic view of semantics of presupposition triggers, thanks to a lexical database comprising 19,500 entries. Secondly, the database was used to investigate a corpus of chat conversations including about 200,000 tokens. The results show that triggers occur mainly as non-informative, maintaining an information already known by all participants of the communication; but, depending on their different features, some of them are systematically associated to a function of anaphora and textual cohesion; others to strengthen social conventions and stereotypes. The informative function, although in a minority proportion, is quantitatively significant only in correspondence to a single class of presupposition triggers.

**Keywords:** Presupposition trigger, Informativeness, Anaphora, Chat

## 1. Introduction

This article presents the results of a study on presupposition in the Italian language, based on a database of about 19,500 lexical presupposition triggers (henceforth: "DB") and on a corpus of computer-mediated communication, specifically formed by chat conversations (henceforth: "the Corpus") including about 200,000 tokens. The research had these phases:

- we developed the DB;
- using the DB, we built semantic classes of triggers based on the features of the presupposed content;
- we used the DB as an automatic tool for text mining, searching all occurrences of lexical presupposition triggers in the Corpus:
- we marked their textual and pragmatic functions within the chat conversations;
- we crossed the previous results, building a multi-level annotation of the semantic and pragmatic phenomena to which the trigger occurrences are associated, finding some profiles that appear in a quantitatively significant proportion.

In the last phases, our attention has been addressed also to the informative uses of presupposition, namely those in which a writer transmits a content not unanimously accepted as true by the audience. The importance of these uses has been highlighted by many authors, recently by Bonyadi, Samuel (2011) and Zare, Abbaspour, Rajaee (2012), who investigated the issue in newspaper editorials, that is, in an unidirectional context; in this case, the aim was to anchor the analysis in a strongly dialogic context that enabled us to see the readers' immediate reactions.

In the following, results of each phase are detailed.

## 2. The Database of Italian Presupposition Triggers (DB)

To setup the DB we started from the English existing lists of presupposition triggers: we can mention at least Sellars (1954), Karttunen (1971), Karttunen (1973), Kiparsky, Kiparsky (1971), Sebba (1987), Konig (1991), Levin (1993), Pi (1995), Tomasello (1992), Landau (2000), Bauerle, Reyle, Zimmermann (2010), Jaszczolt, Turner (2003). Then we translated them into Italian and implemented the list through synonyms, opposites and certain types of derivatives and compounds, using De Mauro (1999) and De Mauro (2010). All collected types are reported in Table 1.

Table 1. Distribution of the DB by type of presupposition trigger

| | Number of lexical entries | % in the DB |
|---|---|---|
| Change-of-state items | 13,551 | 69.5 |
| Factives | 131 | 0.7 |
| Non-factives | 40 | 0.2 |
| Focals | 74 | 0.4 |
| Implicatives | 606 | 3.1 |
| Iteratives | 5,028 | 25.8 |
| Serials | 70 | 0.4 |
| Total | 19,500 | 100 |

Considering that one of the largest Italian dictionaries, De Mauro (1999), reported about 150,000 non-technical headwords, our DB represents a significant portion, 13%. This is a first quantitative measure of the impact of presuppositional triggers on the Italian lexicon.

The distribution is highly concentrated on change-of-state and iterative items, but these categories include many rare words; in contrast, other groups such as focals, which are very small in absolute numbers, include words with a very high frequency in texts (for example: Italian *anche*, English *also* or *too*). 80% of the DB consists of words of basic and common vocabulary, with a high potential impact on the analysis of samples of language; the remaining 20% consists of rare or regional words.

Table 2. Distribution of the DB by grammatical category

| | % of the DB |
|---|---|
| Adjective | 27.2 |
| Complex verbal item | 9.5 |
| Noun | 22.9 |
| Verb | 40.0 |
| Adverb | 0.2 |
| Complex adverbial item | 0.2 |
| Total | 100 |

The survey has been expanded to other categories, in addition to verbs traditionally studied in this field: we included complex lexical items such as Italian *andare in bianco*; nouns and adjectives that are derived from the corresponding verbs (for example: Italian *scongelamento* and *scongelato* derived from the verb *scongelare*). The distribution shown in Table 2 confirms the main importance of the class of verbs, but at the same time gives a more complex framework of all grammatical classes that are involved in the phenomenon of presupposition.

## 3. Semantic Classes of Lexical Presupposition Triggers

Thanks to the first phase of the research, which led to the availability of an organic DB, it was possible to develop a complete view on lexical presupposition triggers. This enabled us to further analyze their classification, taking into consideration all their different semantic properties.

The existence of indefinite presupposition triggers, whose content is non-unique, has already been pointed out, in particular for focals such as English *also* or *too*, by Sbisà (2007), Kripke (2009). We then applied a first distinction between triggers with an open presupposed content (Group A) and triggers with a closed presupposed content (Group B), verifying this criterion within the whole DB.

In addition, in Group B, we made a further discrimination between content that is connoted as standard (B1) and content that is non-connoted or connoted as non-standard (B2). This distinction was developed from an analysis of change-of-state items: for many of them, in fact, it seemed that the presupposition, semantically defined, refers to a state of departure considered normal. We can suppose that such content does not need to be justified by a suitable background (either in the previous discourse or in the active context) to be accepted as appropriate by readers or listeners. This is the case of many verbs that indicate change of moods such as English *to get angry* or *to get sad*; other examples are Italian *annoiarsi, bruciarsi, confondere, ipnotizzare, morire, offendere, preoccupare, rompere, riposare* (English *to bore, to burn, to confuse, to hypnotize, to die, to hurt, to worry, to break, to rest*).

The result of our classification is then the taxonomy described in Figure 1.

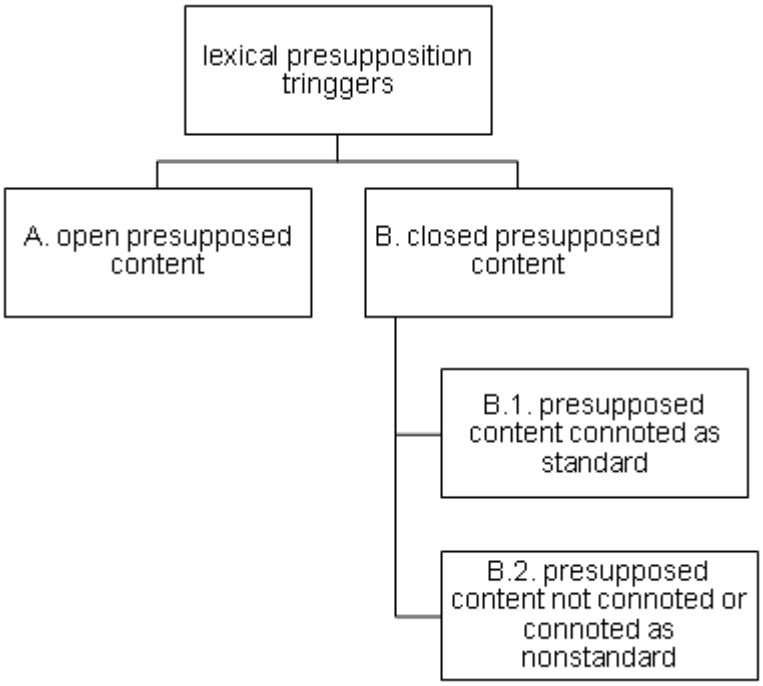


Figure 1. Semantic properties of the presupposed content in the DB.

With the map of Figure 1 we made a semantic annotation of the DB, marking each entry as A, B1 or B2. In this work, we took into account categories of verbal action, such as the opposition between transformative/non-transformative proposed by Bertinetto (2001): in the first case, participants can be found at the end of the process in a state significantly different from the previous one (eg, English *to wake up*, *to stop*); in the second case, participants are involved in a gradual process that does not cause a clear change of their condition (eg, English *to grow*). In this opposition it seemed that non-transformative verbs, in general, have a presupposed content not uniquely defined and so have been included in Group A.

## 4. Textual and Pragmatic Values of the Presupposition Triggers: Results of the Experimental Investigation

For text mining we used the software TaLTaC2 *Trattamento Automatico Lessicale e Testuale per l'Analisi del Contenuto di un Corpus* developed by prof. Bolasco and whose performances are described in Bolasco (2013).

We found 3,291 lexical triggers associated to the merging of a presupposition; for each of them we studied the whole verbal interaction where it was inserted, marking:

1. the presence of elements of the previous discourse that give parallel information to that of the presupposed content, so configuring the presupposition as non-informative and more specifically anaphoric;
2. the presence of elements of the active context (i.e. conditions of the chat session like the entrance or leaving of some participants in the chat) that give parallel information to that of the presupposed content, so configuring the presupposition as non-informative;
3. the presence of readers' reactions, expressing either misunderstanding (such as questions), objection or limitation about the presupposed content (such as negations, specifications), so configuring the presupposition as informative.

What is mentioned in item 1 and 2 represents all contingent elements that may support the presupposed content within the interaction; if they are absent, we may suppose that either the presupposition is informative (as verified by item 3) or is supported by other, non-contingent elements, such as the general encyclopedic knowledge and social conventions of the community.

Figure 2 shows the results of the investigation on all classes of presupposition triggers.

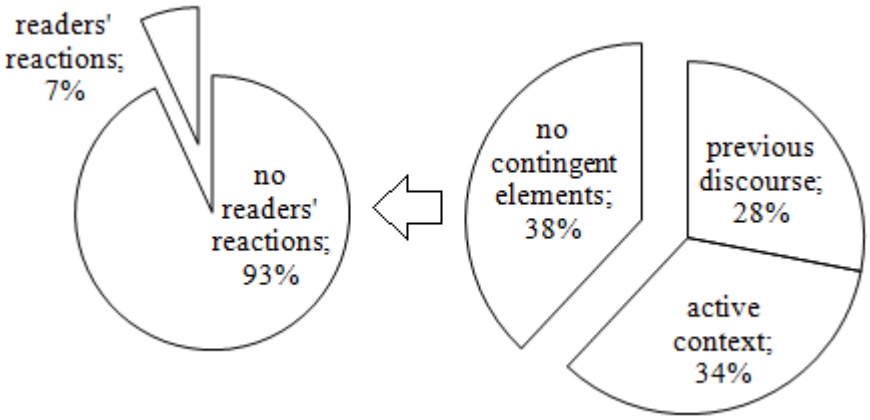


Figure 2. All the presupposition triggers – results of multilevel annotation

In general the use of presupposition triggers is, in the Corpus, non-informative: in fact, in 62% of cases there are contingent elements that complement the information; even in the remaining 38%, there are very few readers' reactions that lead us to assume an informative value of the presupposition.

But, if we detail these findings on each semantic class, the proportions change significantly.

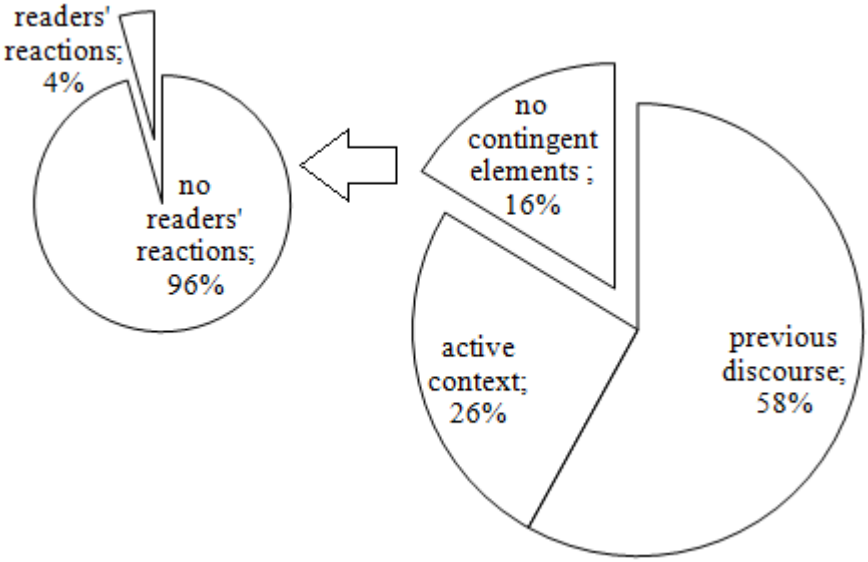


Figure 3. Group A – results of multilevel annotation

Group A, that includes a total of 975 occurrences, shows a clear prevalence of elements of previous discourse (58%) that, added to elements of the active context (26%), reach 84% of the total. This is consistent with the semantic characteristics of this class, that, having an open presupposed content, could otherwise create problems of misunderstanding. The percentage of readers' reactions is smaller than in Figure 2: within the percentage of 16%, only in 4% of cases other chatters feel the need to deny, limit, or ask for clarification on the presupposed content, although there are no contingent elements that provide the information necessary for

its interpretation. Considering all these data, we can conclude that in Group A, while maintaining the overall prevalence of non-informative uses already seen in Figure 2, there is more specifically a prevalence of anaphoric uses.

An example of this profile is the use of Italian *anche*, corresponding to English *also* or *too*.

(1) know|edge: LaR3GinAdiCuoRi si io sono limitato.

[…]

know|edge: e sono anche inetto lo sai?

In (1), the second intervention of "knowledge", using the focal *anche*, implicitly reloads the first intervention, that is quite distant in the course of conversation and consequently needs to be re-activated. In this way, chatters invoke the frame of thematic relevance and thereby facilitate the understanding by other participants.

In Group B1, that includes 441 occurrences of presupposition triggers, there is instead a reversed situation, as shown in Figure 4.

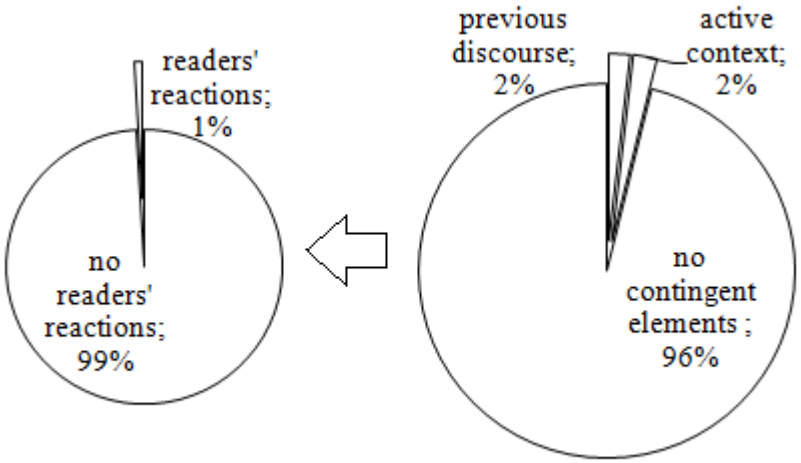


Figure 4. Group B1 – results of multilevel annotation

The presence of contingent elements is much lower than in Group A, reaching 4% of the total. Even in this case, the statistical data is consistent with the semantic characteristics of the class, having a presupposed content connoted as standard that consequently does not need to be integrated or clarified. For the biggest part of the sample, 96%, there are only 1% of readers' reactions. These data suggest that in Group B1 there is a great appeal to the general encyclopedic knowledge and to social conventions.

(2) leggenda: mo offendo miglietto

[…]

miglio: ok

[…]

miglio: tanto io c' ho le spalle larghe

In (2) the use of the change-of-state verb *offendere* (English *to offend*) refers to a state of departure, which is considered normal, of mutual respect and courtesy; the following interaction is playful and "leggenda" purposefully uses words to provoke "miglio". Chatters confirm the presence of social conventions when deliberately challenging them.

Finally, Figure 5 shows the results for group B2, that includes a total of 1,875 occurrences of presupposition triggers.

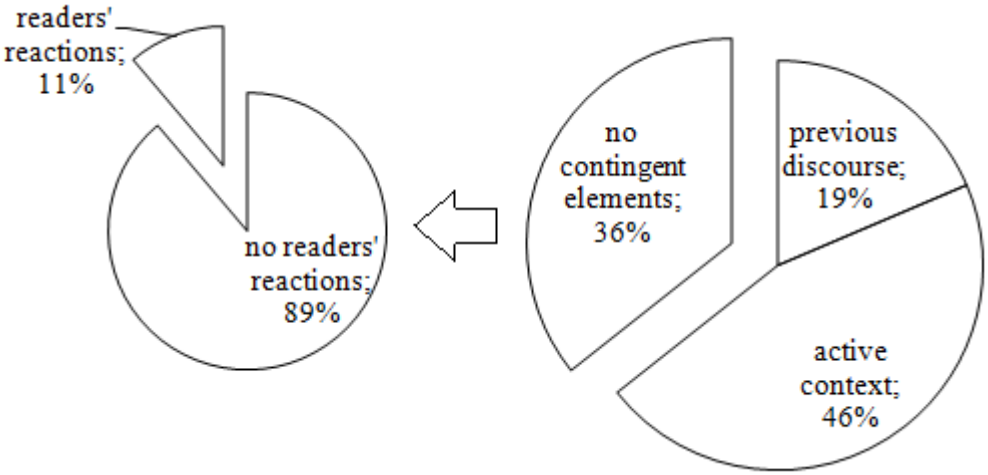


Figure 5. Group B2 – results of multilevel annotation

The main differences compared to Figures 3 and 4 is that in this case the distribution of the percentages is more homogeneous: it is not possible to identify specific trends. The significant number of readers' reactions denotes that the informative function is not as marginal as in groups A and B1.

Reversing the point of view, if we take into account all readers' reactions, we see that they are strongly concentrated in Group B2.

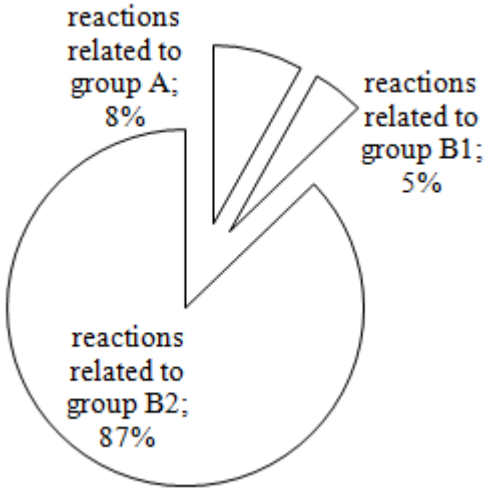


Figure 6. Readers' reactions – distribution by semantic class

Please see example 3.

(3) ba-ba-bambolina: sei tornato in patria?

[…]

Ccnicko79_: E quando sono evaso?

[…]

Ccnicko79_: Nn ricordo

In (3) "bambolina" asks "ccnicko79" if he returned back to his homeland, using the presupposition trigger (iterative) *tornare* (English *to return*); "ccnicko" wonders when he ever left, therefore negating the assumption that he was previously in a different place. In this case the reader reacts about the use of a lexical item of Group B2, by asking a question.

The international state of the art suggests the possibility of establishing systematic linkages between some intrinsic features of the presupposition triggers and functions that they can carry out with greater probability. In particular, an open or non-unique presupposition is supposed to give scarcity of information (and then to have a prevalent anaphoric function): this has been pointed out, for example, for some focal particles such as *also* or *too*, as we already recalled in the beginning of paragraph 3, mentioning Kripke (2009) and Sbisà (2007). Accordingly to this approach, we developed our experimental investigation into a scheme of interrelations between semantic properties of the presupposed content and the textual and pragmatic function of lexical presupposition triggers, based on the above-described statistical data.

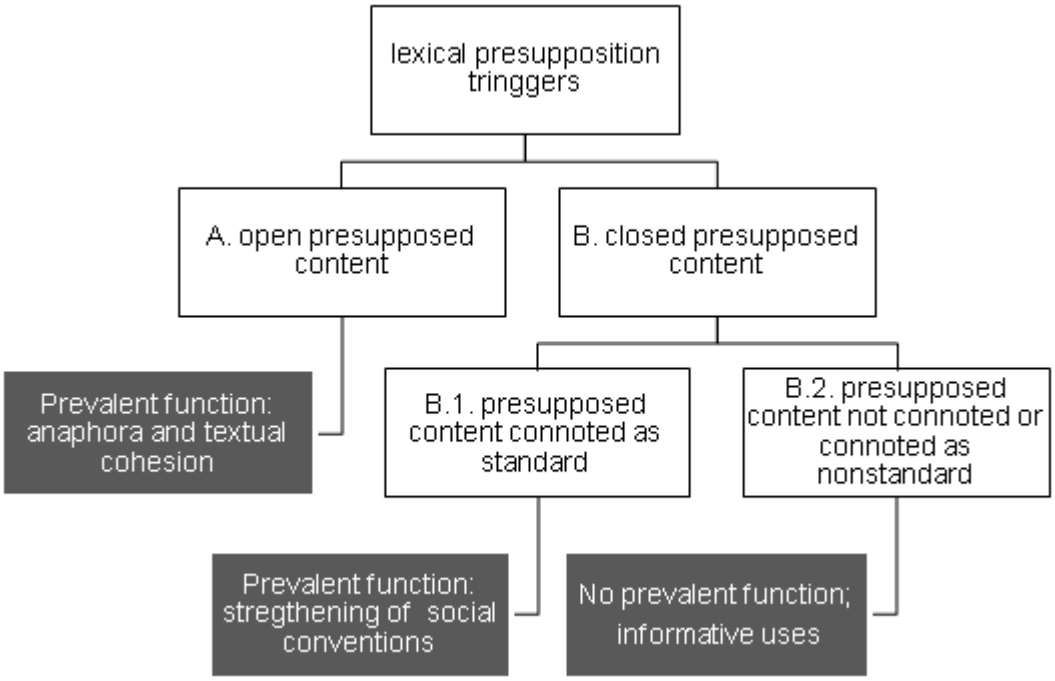


Figure 7. Statistical scheme of interrelations between semantic properties of the presupposed content and the textual and pragmatic function of lexical presupposition triggers

## 5. Conclusion

Group A has a main anaphoric use, therefore having a function of textual cohesion: chatters mostly choose these presupposition triggers to efficiently insert their intervention in the semantic course of the conversation, re-activating previous elements that are useful for interpretation. Group B1 on the other hand seems to strengthen general encyclopedic knowledge and social conventions or stereotypes, playing the pragmatic role of reproducing and maintaining the common field of non-linguistic values and assumptions that are the basis of social relationships. Both prevalent functions of Groups A and B1 appear to be particularly important for communicating via chat: Group A compensates for the difficulties to have a syntactic and semantic continuity, due to the synchronism of written interaction; Group B1 compensates the special conditions of discursive interaction in computer-mediated communication, characterized by temporal proximity and spatial distance, that often give rise to forms of symbolization of social and pragmatic closeness. Group B2 has no specific profile but is the only one where informative function has a quantitatively significant impact.

Further investigations may move from these results to search for more detailed semantic discriminations, especially in Group B2, and for textual specificities associated to the use of each of the groups.

## Acknowledgement

The research is financed by Viterbo University (Italy). Thanks to Rome "La Sapienza" University for providing materials of the Corpus.

## References

Bauerle, R. Reyle, U., & Zimmermann, T. E. (2010). *Presuppositions and Discourse. Essays Offered to Hans Kamp*. Bingley: Emerald.

Bertinetto, P. M. (2001). Il verbo. In L. Renzi, G. Salvi, & A. Cardinaletti (Eds.), *Grande grammatica italiana di consultazione. I sintagmi verbale, aggettivale, avverbiale. La subordinazione* (pp. 13-161). Bologna: Il Mulino.

Bolasco, S. (2013). *L'analisi automatica dei testi. Fare ricerca con il text mining*. Roma: Carocci.

Bonyadi, A., & Samuel, M. (2011). Linguistic Nature of Presupposition in American and Persian Newspaper Editorials. *International Journal of Linguistics, 3*, 1-16. http://dx.doi.org/10.5296/ijl.v3i1.554

De Mauro, T. (1999). *Grande dizionario italiano dell'uso*, Torino: UTET.

De Mauro, T. (2010). *Grande Dizionario Italiano dei Sinonimi e Contrari*. Torino: UTET.

Jaszczolt, K. M., & Turner, K. (2003). *Meaning through language contrast*. Amsterdam: John Benjamins B.V. http://dx.doi.org/10.1075/pbns.100

Karttunen, L. (1971). *The logic of English predicate complement constructions*. Bloomington: Indiana University Linguistics Club.

Karttunen, L. (1973). Presuppositions of Compound Sentences. *Linguistic Inquiry*, 4, 169-193.

Kiparsky, P., & Kiparsky, C. (1971). Fact, in D. Steinberg, & A. Jakobovits (Eds.), *Semantics. An interdisciplinary reader in Philosophy, Linguistics and Psychology* (pp. 345-369). Cambridge: Cambridge University Press.

Konig, E. (1991). *The meaning of focus particles: a Comparative Perspective*. London-New York: Routledge. http://dx.doi.org/10.4324/9780203212288

Kripke, S. A. (2009). Presupposition and Anaphora: Remarks on the Formulation of the Projection Problem, *Linguistic Inquiry*, *40*, 367-386. http://dx.doi.org/10.1093/acprof:oso/9780199730155.003.0012

Landau, I. (2000). *Elements of Control. Structure and Meaning in Infinitival Constructions*. New York: Springer-Verlag. http://dx.doi.org/10.1007/978-94-011-3943-4

Levin, B. (1993). *English verb classes and alternations*. Chicago: Chicago University Press.

Pi, T. (1995). The structure of English Iteratives. In P. Koskinen (Ed.), *Proceedings of the 1995 Annual Conference of the Canadian Linguistic Association* (pp. 435-445). Toronto: Toronto Working Papers in Linguistics.

Sbisà, M. (2007). *Detto non detto. Le forme della comunicazione implicita*. Roma-Bari: Laterza.

Sebba, M. (1987). *The Syntax of serial verbs*. Amsterdam: John Benjamins B.V. http://dx.doi.org/10.1075/cll.2

Sellars, W. (1954). Presupposing. *Philosophical Review, 63,* 197-215. http://dx.doi.org/10.2307/2182346

Tomasello, M. (1992). *First verbs: a case study of early grammatical development*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511527678

Wilson, D. (1975). *Presupposition and Non-truth-conditional Semantics*. New York: New York Academic Press.

Zare, J., Abbaspour, E., & Rajaee N. M. (2012). Presupposition Trigger – A Comparative Analysis of Broadcast News Discourse. *International Journal of Linguistics*, *4*, 734-743. http://dx.doi.org/10.5296/ijl.v4i3.2002

## Copyright Disclaimer