

Measuring the Italian-English lexical gap for action verbs and its impact on translation

Lorenzo Gregori

University of Florence

lorenzo.gregori@unifi.it

Alessandro Panunzi

University of Florence

alessandro.panunzi@unifi.it

Abstract

This paper describes a method to measure the lexical gap of action verbs in Italian and English by using the IMAGACT ontology of action. The fine-grained categorization of action concepts of the data source allowed to have wide overview of the relation between concepts in the two languages. The calculated lexical gap for both English and Italian is about 30% of the action concepts, much higher than previous results. Beyond this general numbers a deeper analysis has been performed in order to evaluate the impact that lexical gaps can have on translation. In particular a distinction has been made between the cases in which the presence of a lexical gap affects translation correctness and completeness at a semantic level. The results highlight a high percentage of concepts that can be considered hard to translate (about 18% from English to Italian and 20% from Italian to English) and confirms that action verbs are a critical lexical class for translation tasks.

1 Introduction

Lexical gap is a well known phenomenon in linguistics and its identification allows to discover some relevant features related to the semantic categorization operated by languages. A lexical gap corresponds to a lack of lexicalization of a certain concept in a given language. This phenomenon traditionally emerged from the analysis of a single language by means of the detection of empty spaces in a lexical matrix (see the seminal works by Leher (1974) and Lyons (1977); see also Kjellmer (2003)). Anyway, lexical gap becomes a major issue when comparing two or more

languages, as in translation tasks (Ivir, 1977). In this latter case, a lexical gap can be defined as the absence of direct lexeme in one language while comparing two languages during translation (Cvilikaitė, 2006). The presence of lexical gaps between two languages is more than a theoretical problem, having a strong impact in several related fields: lexicographers need to deal with lexical gaps in the compilation of bilingual dictionaries (Gouws, 2002); in knowledge representation the creation of multilanguage linguistic resources require a strategy to cope with the lack of concepts (Janssen, 2004); the lexical transfer process is affected by the presence of lexical gaps in automatic translation system, reducing their accuracy (Santos, 1990).

Even if in literature it's possible to find many examples of gaps, it's hard to estimate them. This is due to the fact that most of the gaps are related to small semantic differences that are hard to identify: available linguistic resources usually represent a coarse-grained semantics, so while they are useful to discriminate the prominent senses of words, they can't capture small semantic shifts. In addition to it, a multilanguage resource is required for this purpose, but these resources are normally built up through a mapping between two or more monolingual resources and this cause an approximation in concept definitions: similar concepts tend to be grouped together in a unitary concept that represent the core-meaning and lose their semantic specificities.

2 IMAGACT

Verbs are a critical lexical class for disambiguation and translation tasks: they are much more polysemous than nouns and, moreover, their ambiguity is hard to resolve (Fellbaum et al., 2001). In particular the representation of word senses as sepa-

rate entities is tricky, since their boundaries are often vague causing the senses to be under-specified and overlapping. From this point of view the subclass of general verbs represent a crucial point, because these verbs are characterized by both high frequency in the use of language and high ambiguity.

IMAGACT¹ is a visual ontology of action that provides a video-based translation and disambiguation framework for general verbs. The resource is built on an ontology containing a fine-grained categorization of action concepts, each represented by one or more video prototypes as recorded scenes and 3D animations.

IMAGACT currently contains 1,010 scenes which encompass the action concepts most commonly referred to in everyday language usage. Data are derived from the manual annotation of verb occurrences in spontaneous spoken corpora (Moneglia et al., 2012); the dataset has been compiled by selecting action verbs with the highest frequency in the corpora and comprises 522 Italian and 554 English lemmas. Although the set of retrieved actions is necessarily incomplete, this methodology ensures to have a significant picture of the main action types performed in everyday life².

The links between verbs and video scenes are based on the co-referentiality of different verbs with respect to the action expressed by a scene (i.e. different verbs can describe the same action, visualized in the scene). The visual representations convey the action information in a cross-linguistic environment and IMAGACT may thus be exploited to discover how the actions are lexicalized in different languages.

In addition to it IMAGACT contains a semantic classification of each lemma, that is divided into Types: each verb Type identifies an action concept and contains one or more scenes, that work as prototypes of that concept. Type classification is manually performed in Italian and English in parallel, through a corpus-based annotation procedure by native language annotators (Moneglia et al., 2012); this allowed to have a discrimination of verb Types based only on the annotator competence, without any attempt to fit the data into predefined semantic models. Validation results (Gagliardi, 2014) highlight a good rate of

¹<http://www.imagact.it>

²see Moneglia et al. (2012) and Moneglia (2014b) for details about corpora annotation numbers and methodology.

Type discrimination agreement: a Cohen k of 0,82 for 2 expert annotators and a Fleiss k of 0.73 for 4 non-expert ones³.

For these features IMAGACT ontology is a reliable data source to measure the lexical gap between Italian and English: in fact verb Types are defined independently, but linked together through the scenes. The comparison of Types in different language through their action prototypes allows to identify the action concepts that are shared between the two languages and the ones that don't match with any concept in the other language; in this case we have a lexical gap.

3 Type relations

In this frame we can perform a set-based comparison, considering a Type as just a set of scenes. A Type is a lexicalized concept, so a partition of the meaning, but semantic features are not represented in the ontology and, in fact, they are unknown: data are derived from the ability of competent speaker in performing a categorization of similar items with respect to a lemma, without any attempt to formalize senses. So if we look at the database we can say that Types are merely sets of scenes.

Comparing a Type (T_1) of a verb in source language (V_1) with a Type (T_2) of a verb in target language (V_2) we can have 5 possible configurations:

1. $T_1 \equiv T_2$: two Types are equivalent if they contain the same set of scenes;
2. $T_1 \cap T_2 = \emptyset$: two Types are disjoint if they don't share any scene;
3. $T_1 \subset T_2$: T_1 is a subset of T_2 if any scene of T_1 is also a scene of T_2 and the 2 Types are not equivalent;
4. $T_1 \supset T_2$: T_1 is a superset of T_2 if any scene of T_2 is also a scene of T_1 and the 2 Types are not equivalent;
5. $T_1 \cap T_2 \neq \emptyset \wedge T_1 \not\subseteq T_2 \wedge T_1 \not\supseteq T_2$: two Types are partially overlapping if they share some scenes and each Type have some scenes not belonging to the other one.

³Inter-annotator agreement (ITA) measured on WordNet by expert annotators on comparable verb sets (score for fine-grained sense inventories): ITA = 72% on SemEval-2007 dataset (Pradhan et al., 2007); ITA = 71% on SensEval-2 dataset (Palmer et al., 2007).

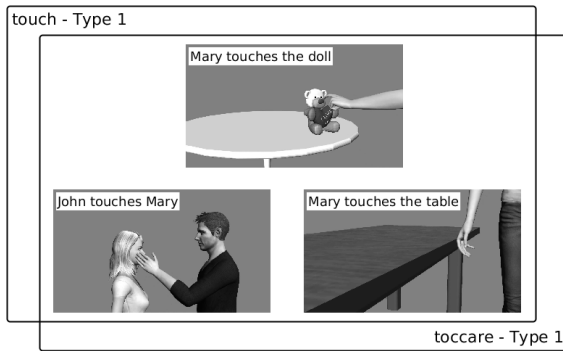


Figure 1: Two Equivalent Types belonging to the Italian verb *toccare* and to the English verb *touch*.

It's important to discuss these cases separately, because each one of them highlights a different semantic relation between verbs and has different implications for translation.

When two Types are equivalents (case 1) the 2 languages share the action concept the Types represent: we could say that there is an interlinguistic concept. This case is not problematic for translation: each occurrence of the verb V_1 that belongs to Type T_1 can be translated with V_2 ; moreover we can apply V_1 to translate any occurrence of V_2 belonging to T_2 .

For example one Type of the English verb *to touch* and one Type of the Italian verb *toccare* are equivalent. They share 3 video scenes: *Mary touches the doll*, *Mary touches the table* and *John touches Mary* (see Fig. 1). Each scene is connected to a different set of verbs (i.e. *to brush*, *to graze*, *to caress*), representing a specific semantic concept, but they are kept together by a more general concept both in Italian and in English. So in any of these actions the verb *to touch* can be safely translated in Italian with *toccare* and vice versa.

If two Types are disjoint (case 2) the Types refer to unrelated semantic concepts and we can assume that translation between an occurrence of V_1 belonging to T_2 can not be translated with V_2 .

In cases 3 and 4 the Types are hierarchically related and we can assume the existence of a semantic relation that links a general Type with a specific one. Although we can not induce the type of this relation that could be *hyponym*, *entailment*, *troponym* and so on. In this configuration we can see that translation is safe from specific to general, but not vice versa: in case 3 any occurrence of V_1 belonging to T_1 can be translated with V_2 , while in case 4 V_2 can not be safely applied, because it

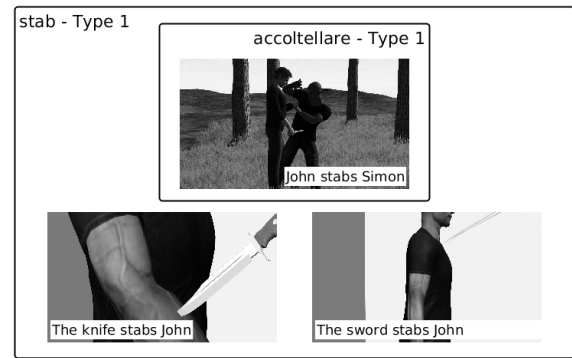


Figure 2: Two Hierarchically related Types belonging to the Italian verb *accoltellare* and to the English verb *stab*.

encodes only a sub-part of the concept represented by T_1 .

For example Type 1 of the English verb *to stab* and Type 1 of the Italian verb *accoltellare* categorize action where a sharp object pierces a body, but while *stab* can be applied to describe actions independently on their aim and the tool used, *accoltellare* is applicable only when the agent voluntarily injures someone and the action is accomplished with a knife. In this case the Italian Type is more specific than the English one, so translation is safe from Italian to English (*stab* can be used to translate any occurrence of *accoltellare-Type 1*), but not vice versa: *stab-Type 1* can not be always translated with *accoltellare*, because a part of its variation is covered by other Italian verbs like *trafiggere*, *penetrare* or *attraversare*.

Finally a partial overlap between Types (case 5) doesn't allow to induce any semantic relation between Types: in these cases we have different concepts that can refer the same action. Normally these happen when the action is interpreted from two different points of view and categorized within unrelated lexical concepts. In this case we have a translation relation between V_1 and V_2 without having any semantic relation between their Types.

For example the Italian verb *abbassare*, that is frequently translated with *lower* in English, can also be translated with *position* when applied to some (but not every) actions belonging to Type 1, categorizing actions involving the body; moreover we have the same translation relation from English to Italian where sometimes (but not always) *position-Type 2* can be translated with *abbassare*. Here there are two Types that represent semanti-

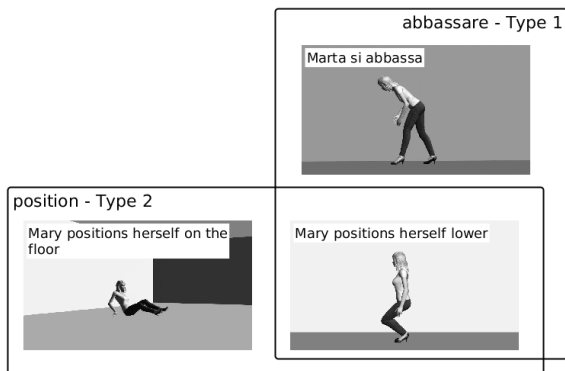


Figure 3: Two partially overlapping Types belonging to the Italian verb *abbassare* and to the English verb *position*.

cally independent concepts, but that can both be applied to describe some actions, like *Mary positions herself lower* and other similar ones.

This happens rarely in Italian - English (14 Types on our dataset) and in any of these cases there are other translation verbs as possible alternatives. Despite this, Type overlaps identification is very relevant, because it allows to discover unexpected translation candidates (i.e. target verbs that have a translation relation but not a semantic relation with the source verb) that can not be extracted from a lexico-semantic resource. In addition to it Type overlaps identification is crucial if the target verb is the only one translation possibility and this can happen, especially between two languages that are very far: some evidences for example have been discovered in Italian and Chinese (Pan, 2016) through a deep comparison of Italian Types with Chinese verbs that refer to the same scenes. This work allowed to identify some positive occurrences of this interesting phenomenon, but can not be exploited for its numeric quantification: indeed an exhaustive analysis that involves the relation between action concepts can be made only between Italian and English, since IMAGACT contains the verb Type discrimination in these two languages only.

4 Lexical gap identification

4.1 Dataset building

In order to measure the lexical gaps in Italian and English we created a working dataset by selecting the set of Types that have a full mapping in the two languages. We need to consider that IMAGACT annotation process has been carried out in

several steps: firstly verbs were annotated through a corpus-based procedure and Types were created and validated by mother tongue speakers on the basis of their linguistic competence; then for each concept a scene was produced to provide a prototypical representation of it; after that a mapping between Italian and English was performed by linking the scenes to the Types of each language; finally annotators were requested to recheck each scene and add the missing verbs that are applicable to it. This last revision enriched the scene with more verbs that don't belong to any Type.

We decided to exclude from the dataset all the scenes (and the related Types) that contain *untyped* verbs, considering that a partial typing does not ensure the coherence of verb Type discrimination: in fact it's not possible to be sure that the creation of Types for these new instances would preserve the original Type distinction.

After this pruning we obtained a set of 1,000 Italian Types and 1,027 English Types, that refer to 501 and 535 verbs respectively (see Table 1).

	IT	EN
Types	1,000	1,027
Verbs	501	535
Scenes	980	917

Table 1: Number of Types, verbs and scenes belonging to the dataset.

4.2 Methodology

According to our dataset, we can easily estimate the lexical gap by measuring the number of Types in source language that don't have an equivalent Type in target language. Namely for each concept in source language we are going to verify if there is a concept in target language that refer to the same set of actions (represented by video prototypes); if the match is not found we have a lexical gap in target language.

As we can see in table 2, the action concepts that are lexicalized in Italian and without a corresponding match in English are 33,6% (English gap); on the contrary the Italian gap for English concepts is 29,02%.

Before going ahead we need to do some considerations about these numbers. First of all we can see that these percentages are much higher than the ones calculated by Bentivogli and Pianta (2000), that found 7,4% of gaps for verbs in

	IT → EN	EN → IT
Total Types	1,000	1,027
Equiv. Types	664 (66,4%)	729 (70,98%)
Lexical gaps	336 (33,6%)	298 (29,02%)

Table 2: Types in source language that have and have not an equivalent Type in target language.

English-to-Italian comparison. This is a big shift, but it’s not surprising if we consider the differences of the two experiments in terms of methodology and dataset:

- IMAGACT Type distinction is more fine-grained in respect to WordNet synsets (Bartolini et al., 2014);
- the experiment by Bentivogli and Pianta was led on MultiWordNet, in which multilingual Wordnets are created on the basis of the Princeton Wordnet sense distinction (Pianta et al., 2002); this methodology introduce an approximation in the concepts definition;
- the 7,4% of Bentivogli and Pianta is a general value on verbs, while our experiment is focused on action verbs, which are a strongly ambiguous lexical class (Moneglia, 2014a);
- the dictionary-based methodology proposed by Bentivogli and Pianta is nearly opposite to IMAGACT reference-based approach.

Beyond these general considerations a lemma-by-lemma comparison with the experiment of Bentivogli and Pianta (whose dataset is currently not available) would better explain this numeric difference.

5 Lexical gaps and translation problems

Besides a general measure of the gaps for action concept it’s important to go a step beyond to verify in which cases the presence of a lexical gap impacts the translation quality. In order to do this, we divided the Types without an equivalent in target language in three categories:

- *leaf Types*: these Types in source language represent concepts that are more specific than other ones in target language; in this case the only Type in target language that have a partial match with the Type in source language is a superset (case 3);

- *root Types*: these Types in source language represent concepts that are more general than other ones in target language: the only Type in target language that have a partial match with the Type in source language is a subset (case 4);
- *middle Types*: these Types have a partial match in target language both with a more general Type and with a more specific one (both cases 3 and 4).

As we mentioned before we did not find any case in which a partial overlapping Type (case 5) is the only one possible match in Italian and English comparison; so these cases are counted within the three categories above.

5.1 Root Types and uncertain translations

Starting from this classification we can see that *root Types* are the critical ones in terms of translation: in fact we don’t have a unique lexicalized concept in target language that is able to represent the concept in source language; instead we have more than one Type (and multiple verbs) that cover different subparts of the whole general concept variation. In these cases we need to have extra information about the action in order to translate it properly. From a computational point of view we can say that a word sense disambiguation of the source verb is not enough to reach a correct translation verb.

The two sentences *The cell phone lands on the carpet* and *The pole vaulter lands on the mat*, for example, belong to the same action concept according to the semantics of the verb *to land*⁴. In Italian there is not a unique Type that collects these two actions: it’s possible to use *atterrare* for the athlete, but it is not allowed for the phone, for which we need to make a semantic shift and use the verb *cadere* (that is more similar to *fall down*). Again *cadere* is not appropriate for the athlete, because it implies that the athlete stumbles and falls.

So this action concept that is lexicalized in English with *to land* does not have a unique translation verb in Italian, and extra informations are required to translate it properly (if the theme is an human being or an object, in this specific case).

Table 3 show the number of *leaf*, *root* and *middle Types* in Italian and English; we can see that

⁴unlike *The butterfly lands on the flower* or *The airplane lands* that belong to different concepts of *to land*.

	IT → EN	EN → IT
Lexical gaps	336	298
Leaf Types	217 (64.58%)	200 (67.58%)
Root Types	47 (13.99%)	43 (14.43%)
Middle Types	72 (21.43%)	55 (18.46%)

Table 3: Number of Leaf, Root and Middle Types in Italian and English (percentages on the lexical gaps).

root Types represent the 14% of the lexical gaps in both the languages, corresponding to 4-5% of the total Types.

5.2 General Types and lossy translations

Root Types are the most critical case for a translation task, because they affect the correctness; besides there are also other kinds of lexical gaps that impact on translation. In particular is useful to estimate how semantically far is the best translation candidate in the cases in which we can apply a more general Type to translate the concept in the source language. In fact in both *leaf* and *middle Types* we have a Type in target language that is more general to the source one, so it is safely applicable to any occurrence belonging to the source Type. This is not free from problems, because in translation we use a more general verb, so we miss some semes that are encoded in the source verb. In fact in this case we still have a translation problem, which is not in finding a possible target verb, but in adding more information in other lexical element of the sentence to fill the lack of semantic information. In this case the gap does not affect the correctness of the translation, but its completeness.

For example the English verb *to plonk* does not have a correspondence in Italian. In particular a sentence like *John plonks the books on the table* belongs to a Type of *plonk* that is a *leaf Type* (so there is a possible translation verb in Italian), but for which the nearest Italian Type is much wider, belonging to the very general verb *mettere*. In this case it’s possible to translate in Italian with *John mette i libri sul tavolo*, but losing all the information regarding the way the books are placed on the table (*mettere* is more similar with *to put*); an addition of other lexical elements to the sentence is required to fill this gap in Italian.

Conversely we can say that a small distance between the source and the target Type does not have

a negative effect on translation. Type 1 of the English verb *to throw* and Type 1 of the Italian verb *lanciare* categorize a wide set of actions in which an object is thrown by a person independently on the presence of a destination or on the action aim (*John throws the bowling bowl*, *John throws the rock in the field*, *John throws the paper in the box* etc.). However these two Types are not equivalent, because the Italian one comprise also actions performed in a limited space with a highly controlled movement, like *Marco lancia una moneta*, that require another verb in English like *to toss* (*Marco tosses a coin*). In this case the small gap between the Italian concept and English one does not affect the translation: in fact we can say that *lanciare* can be used to translate properly any action belonging to *throw - Type 1*.

Given this consideration a measure of the semantic distance with the translation verb is useful to evaluate the loss: this can be easily done from IMAGACT dataset by calculating the ratio between the cardinality (i.e. the number of scenes) of the source Type, T_1 , and the one of the nearest target Type, T_2 (the Type with the minimum cardinality among the Types in target language that are supersets of the source Type). This ratio estimates the overlapping between the Types:

$$overlap = \frac{|T_1|}{|T_2|}$$

Data are represented in Figure 4, reporting the number of Types (Italian and English) for each overlap values, where this values are divided in 10 ranges.

We considered semantically distant those Types with $overlap < 0.4$ (sharing less than 2 scenes over 5). These *high distance Types* (see Table 4) are 150 for Italian (51.9% of *leaf + middle Types* and 15% of the total Types) and 145 for English (56.86% of *leaf + middle Types* and 14.12% of the total Types).

Basically we see that not only *root Types*, but also a relevant part of *leaf* and *middle Types* (more than 50% both in Italian and English) represent a critical point for translation.

	IT → EN	EN → IT
Leaf+Mid T.	289	255
Low dist. T.	139 (48.1%)	110 (43.14%)
High dist. T.	150 (51.9%)	145 (56.86%)

Table 4: Distance from the nearest general Type in target language.

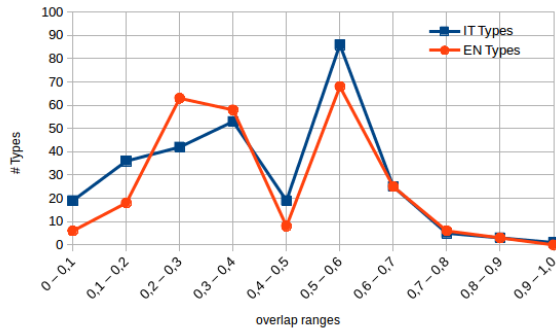


Figure 4: Number of Italian and English Types for each overlap range.

Within this numbers, that are quite homogeneous between the two languages, we can see that in the overlap range from 0 to 0.2 there are much more Italian Types than English ones (19% of *leaf + middle Types* against 9%); conversely English Types are more distributed in the range from 0.3 to 0.4 (Figure 4). This means that in this area of extreme distance between the source and the target concept, we have an higher semantic loss in the translation from Italian to English.

Finally we can have an overall value of translation critical Types, by summing up the ones belonging to *high distance Types* class and the *root Types*. The verbs these Types belong to are the verbs for which the selection of a good translation candidate is problematic. Results are reported in Tables 5 and 6 and confirm that lexical gaps in action verbs have a strong impact on translation.

	IT	EN
Total Types	1,000	1,027
Root Types	47 (4.7%)	43 (4.2%)
High dist. T.	150 (15.0%)	145 (14.12%)
Critical Types	197 (19.7%)	188 (18.3%)

Table 5: Number of translation critical Types.

	IT	EN
Total Verbs	501	535
Verbs w/ r.T.	39 (7.78%)	38 (7.1%)
Verbs w/ h.d.T.	109 (21.76%)	125 (23.36%)
Critical Verbs	136 (27.15%)	154 (28.79%)

Table 6: Number of verbs with *root Types* and *high distance Types*.

6 Conclusions

In this paper a methodology for measuring the lexical gap of action verbs is described and applied to Italian and English, by exploiting IMAGACT ontology. We measured 33.6% of English gap and 29.02% of Italian gap. Then this result have been investigated, in order to discover when and why a lexical gap can affect a translation task. The results show that 19.7% of Italian Types and 18.3% of English ones represent action concept that are critical from a translation perspective: these concepts are lexicalized by 27.15% of the Italian verbs and 28.79% of the English verbs that we considered in our analysis. In addition to it the distinction between concepts that can not be correctly translated with a single lemma (*root Types*) and concepts that can be translated with a sensible semantic loss (*high distance Types*) is a relevant information that can lead to a different translation strategy.

Finally we feel important to note that behind these numeric values there are lists of verbs and concepts and this information could be integrated in Machine Translation and Computer Assisted Translation Systems to improve their accuracy.

Acknowledgments

This research has been supported by the MODELACT Project, funded by the Futuro in Ricerca 2012 programme (Project Code RBFR12C608); <http://modelact.lablita.it>.

References

- Roberto Bartolini, Valeria Quochi, Irene De Felice, Irene Russo, and Monica Monachini. 2014. From synsets to videos: Enriching italwordnet multimodally. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may. European Language Resources Association (ELRA).
- Luisa Bentivogli and Emanuele Pianta. 2000. Looking for lexical gaps. In *Proceedings of the ninth EU-RALEX International Congress*, Stuttgart, Germany.
- Jurgita Cvilikaitė. 2006. Lexical gaps: resolution by functionally complete units of translation. *Darbai ir dienos*, 2006, nr. 45, p. 127-142.
- Christiane Fellbaum, Martha Palmer, Hoa Trang Dang, Lauren Delfs, and Susanne Wolf. 2001. Manual and

- automatic semantic annotation with wordnet. *WordNet and Other Lexical Resources*, pages 3–10.
- Gloria Gagliardi. 2014. *Validazione dell'ontologia della IMAGACT per lo studio e la diagnosi del Mild Cognitive Impairment*. Ph.D. thesis, University of Florence.
- Rufus Hjalmar Gouws. 2002. Equivalent relations, context and cotext in bilingual dictionaries. *Hermes*, 28(1):195–209.
- Vladimir Ivir. 1977. Lexical gaps: A contrastive view. *Studia Romanica et Anglica Zagrabiensia*, (43):167–176.
- M. Jansseen. 2004. Multilingual lexical databases, lexical gaps, and simulda. *International Journal of Lexicography*, 17(2):137–154.
- Gran Kjellmer. 2003. Lexical gaps. In *Extending the scope of corpus-based research*, pages 149–158. Brill.
- A. Lehrer. 1974. *Semantic Fields and Lexical Structure*. North-Holland linguistic series, 11. North-Holland.
- John Lyons. 1977. Semantics, volume i. *Cambridge UP, Cambridge*.
- Massimo Moneglia, Francesca Frontini, Gloria Gagliardi, Irene Russo, Alessandro Panunzi, and Monica Monachini. 2012. Imagact: deriving an action ontology from spoken corpora. *Proceedings of the Eighth Joint ACL-ISO Workshop on Interoperable Semantic Annotation (isa-8)*, pages 42–47.
- Massimo Moneglia. 2014a. Natural Language Ontology of Action: A Gap with Huge Consequences for Natural Language Understanding and Machine Translation. In Zygmunt Vetulani and Joseph Mariani, editors, *Human Language Technology Challenges for Computer Science and Linguistics*, volume 8387 of *Lecture Notes in Computer Science*, pages 379–395. Springer International Publishing.
- Massimo Moneglia. 2014b. The semantic variation of action verbs in multilingual spontaneous speech corpora. In T. Raso and H. Mello, editors, *Spoken Corpora and Linguistics Studies*, pages 152–190. John Benjamins Publishing Company.
- Martha Palmer, Hoa Trang Dang, and Christiane Fellbaum. 2007. Making fine-grained and coarse-grained sense distinctions, both manually and automatically. *Natural Language Engineering*, 13(02):137–163.
- Yi Pan. 2016. *Verbi dazione in Italiano e in Cinese Mandarino. Implementazione e validazione del cinese nell'ontologia interlinguistica dell'azione IMAGACT*. Ph.D. thesis, Università degli Studi di Firenze.
- Emanuele Pianta, Luisa Bentivogli, and Christian Girardi. 2002. MultiWordNet: developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet*, pages 21–25.
- Sameer S. Pradhan, Edward Loper, Dmitriy Dligach, and Martha Palmer. 2007. Semeval-2007 task 17: English lexical sample, srl and all words. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 87–92. Association for Computational Linguistics.
- Diana Santos. 1990. Lexical gaps and idioms in machine translation. In *Proceedings of the 13th Conference on Computational Linguistics - Volume 2, COLING '90*, pages 330–335, Stroudsburg, PA, USA. Association for Computational Linguistics.