

Tracing Metonymic Relations in T-PAS: An Annotation Exercise on a Corpus-based Resource for Italian

Emma Romani
University of Pavia
Department of Humanities
emma.romani01@
universitadipavia.it

Elisabetta Ježek
University of Pavia
Department of Humanities
jezek@unipv.it

Abstract

In this paper we address the main issues and results of a research thesis (Romani, 2020) dedicated to the annotation of metonymies in T-PAS, a corpus-based digital repository of Italian verbal patterns (Ježek et al., 2014). The annotation was performed on the corpus instances of a selected list of 30 verbs and was aimed at both implementing the resource with metonymic patterns and identifying and creating a map of the metonymic relations that occur in the verbal patterns. The annotated corpus data (consisting of 1218 corpus instances), the patterns, and the relations can be useful for NLP tasks such as metonymy recognition.¹

1 Introduction

Metonymy is a language phenomenon for which one referent is used to denote another referent associated with it (Lakoff & Johnson, 1980; Fauconnier, 1985; Ježek, 2016). For example, in the sentence ‘he drank a glass at the pub’, *glass* (the *metonymic* or *source type* denoting a container) refers to its content (the *target type*, a beverage). In our research, we investigated metonymy from a corpus-based perspective, through the analysis of corpus data and the annotation performed in T-PAS, a corpus-based resource for Italian verbs. T-PAS consists of a repository of typed predicate argument structures (Ježek et al., 2014), i.e. verbal patterns together with semantically-specified arguments, linked to manually annotated corpus instances (see Section 3.1). An example of a pattern (or *t-pas*) for the verb

bere ‘to drink’ is reported in Figure 1:

```
1 [Animate] bere [Beverage]
  [Animate] ingerisce, assume [Beverage]
```

Figure 1. Pattern 1 of the verb *bere* (‘to drink’) in T-PAS together with its sense description

where [Animate] and [Beverage] are the semantic types specifying the subject and object positions.

The annotation of metonymies was performed on the corpus instances of a list of 30 verbs contained in T-PAS (taken from Ježek & Quochi, 2010). As emerged from this background study, the semantic properties of those verbs were likely to convey metonymies in their argument structures. Starting from this list, our work was intended as an implementation of the resource; specifically, we annotated metonymic corpus instances and created metonymic sub-patterns linked to them.

The research had several aims. First, we were interested in studying qualitatively the phenomenon in and through the corpus instances and in implementing the annotation tool of the resource with a specific feature that allowed us to encode metonymic arguments in the verbal patterns. For the latter purpose, we collaborated with the Faculty of Informatics at Masaryk University of Brno (CZ); they gave us the technical support for the implementation of the annotation tool.

Second, our intention was to conceive a general theoretical framework to represent the metonymies found through the qualitative corpus analysis, by designing a map of metonymies and by drafting a list of the metonymic *relations* that occur in the verbal patterns (see Section 4).

The paper is organized as follows. In Section 2 we present related studies. In Section 3 we describe the methodology we followed in annotating the corpus instances for metonymies, together with a brief introduction to T-PAS. In Section 4 we present the results of our annotation: a generalization of the

¹ Copyright ©2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

metonymic relations found, and a map which visually highlights the semantic and cognitive connections between the semantic types. Further developments of the project are described in Section 5; our intention is to enrich the number of analysed verbs and eventually add new types of metonymic relations.

2 Related works

Corpus-based studies on metonymy are often intended for NLP tasks. Markert & Nissim (2006), provide a corpus-based annotation scheme for metonymies with the aim of improving automatic metonymy recognition and resolution. Related to it, Markert and Nissim (2007) present the results of a supervised task on metonymy resolution; an analogous task has been addressed by Pustejovsky et al. (2010) within the scope of SemEval-2010. A recent study elaborated a computational model based on the dataset of Pustejovsky et al. (2010) for the detection of metonymies (McGregor et al., 2017).

Corpus-based studies on metonymies do not necessarily address NLP tasks. An attempt to implement corpus-based resources to display metonymies is described in Ježek & Frontini (2010). Also, Pustejovsky & Ježek (2008) present a corpus investigation aimed at identifying metonymic mechanisms in predicate-argument constructions from a theoretical perspective. Finally, Marini & Ježek (2020) performed an equivalent corpus-based metonymy annotation on a sample of 101 Croatian verbs within the scope of CROATPAS (Marini & Ježek, 2019), sister project of T-PAS.

3 Resource and methodology

3.1 The resource: T-PAS

T-PAS is the corpus-based resource used in this research. It consists of a repository of Typed Predicate Argument Structures (T-PAS) (Ježek et al., 2014) for Italian verbs. The resource consists of three components:

- 1) a repository of corpus-derived predicate argument structures for verbs with semantic specification of the arguments, linked to lexical units (verbs);
- 2) an inventory of about 200 corpus-derived semantic classes for nouns, relevant for the disambiguation of the verb in context;
- 3) a corpus² of sentences that instantiate T-

² The corpus is a reduced and cleaned version of It-WaC (Baroni et al., 2009), a corpus of Italian texts, available in the Sketch Engine tool (Kilgarriff et al., 2014).

PAS, tagged with lexical unit (verb) and pattern number.

Typed predicate argument structures (or *t-pass*) are patterns which display the syntactic and semantic properties of verbs: for each meaning of a verb a specific t-pass is provided. Verb sense is determined by the arguments it combines with (subject, object, etc.), which are defined through a specific Semantic Type.³

T-pass are corpus-derived: patterns were acquired through the manual clustering and annotation of corpus instances for each verb following the CPA procedure (Hanks, 2013). Each t-pass in the resource is labelled with a number and connected to a list of corpus instances, realizing the specific verb meaning. Each pattern is associated with a *sense description*, a brief definition of the meaning of the verb (see the second line in Figure 1). Each pattern can have *sub-patterns* created by annotators, for corpus instances that do not reflect the prototypical semantic behaviour of the verb or of its arguments, as in metonymic uses. Like their patterns, sub-patterns are connected to annotated instances from the corpus. In our work, we implemented the annotation tool by adding a new label ('.m'), which we used to annotate metonymic uses in sub-patterns (see Figure 2).

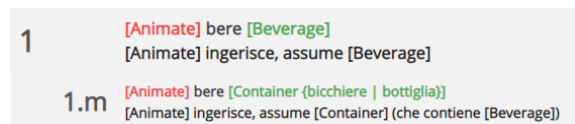


Figure 2. Metonymic sub-pattern for t-pass 1 of the verb *bere* ('to drink') in T-PAS

3.2 Methodology

We conceived an empirical methodology in order to get significant results from the corpus analysis: we manually extracted significant instances from the corpus and annotated them as metonymic instances under their specific pattern. In order to annotate the instances, we exploited the Sketch Engine functions available for analysing the corpus. The annotation scheme can be summarized as follows:

- 1) Random sampling of about 200 corpus instances for each of the 30 verbs (the sample allowed to reduce the time spent in skimming the instances, still providing a balanced overview of the kind of instances contained in the corpus);

³ Semantic Types are expressed through square brackets (e.g. [Animate], [Beverage]) and are organized in a hierarchy, called the System of Semantic Types (see Ježek, 2019 for a more detailed account).

<i>verbs</i>	<i>n. of annotated instances</i>	<i>verbs</i>	<i>n. of annotated instances</i>	<i>verbs</i>	<i>n. of annotated instances</i>
accusare	34	concludere	39	parcheggiare	19
annunciare	40	contattare	27	raggiungere	51
arrivare	47	continuare	21	recarsi	81
ascoltare	103	divorare	0	rimbombare	28
atterrare	76	echeggiare	24	sentire	27
avvisare	22	finire	66	sorseggiare	32
bere	93	informare	16	udire	35
chiamare	16	interrompere	39	venire	63
cominciare	19	leggere	84	versare	12
completare	42	organizzare	13	visitare	49

Table 1. List of the Italian verbs with number of annotated instances in ItWaC corpus

- 2) Manual annotation of the metonymic instances through the sublabel (signalled with “.m”);
- 3) Implementation of the sub-pattern in the resource by adding metonymic semantic types (see 1.m in Figure 1);
- 4) Definition of the metonymic relation (see Table 2) between the source and the target semantic type (e.g. [Container] ‘contains’ [Beverage]), with its encoding in the sense description, translated in Italian (see Figure 2).

In Table 1, we show the number of instances annotated for each of the 30 verbs. Overall, the dataset consists of 1218 annotated instances. The number of instances from the random sample can vary from a verb to another, because verbs have different frequencies in the corpus and metonymic phenomena can be more or less pervasive according to the verb under examination. Some cases (e.g. *divorare* – ‘to devour’ – in Table 1) did not provide any metonymic instance at all (for an explanation and further discussion on this point, see Romani, 2020).⁴

The annotation procedure was conducted manually by one single annotator (the first author) and, so far, it was not possible to evaluate our annotation procedure as we focused on the qualitative analysis and the retrieval of the relations: it is our intent for the future, as it is essential for further progresses in the research.

Currently, the adopted annotation scheme did not provide ambiguous cases, as metonymies were usually clear-cut and the shift of referent from the source to the target semantic type easily identifica-

ble. This may differ from metaphors, for example, where the shift between literal and non-literal meaning may be perceived as more gradual. However, further investigation needs to be done through the annotation of a higher number of instances (expanding the list of verbs) and the comparison and the evaluation of the annotation results of more than one annotator.

4 Results

The overall aim of the research was to give a theoretical account of the metonymic relations found through the corpus analysis and pattern annotation. Therefore, the main results of the study are a list of metonymic relations between the *target* and the *metonymic* (source) *semantic type* (Table 2, Appendix) and a map where the target semantic types are connected to the metonymic types, and the relation between the two is expressed (Figure 3).

The second column in Table 2 (Appendix) reports the 37 *relations* we identified and encoded (the relations are grouped according to their target type, following this order: [Human], [Location], [Document], [Beverage], [Vehicle], [Sound]). The *relation* is a short description that illustrates how the metonymic semantic type is connected to the target semantic type; for example (see the highlighted line of the table), [Container] is the metonymic semantic type (first column) and ‘contains’ is the relation (second column) which links [Container] to the target semantic type [Beverage] (third column).⁵ An instance for this is: ‘we went out to drink *a glass*’ (*glass* ‘contains’ *something to drink*). The fourth column contains an instance in Italian

⁴ In some cases, additional instances were included, if the number of metonymic instances provided by the sample was not sufficient to exemplify a specific relation. Instances with arguments and semantic types analogous to the ones already tagged were selected. To do so, we exploited other Sketch Engine functions (see Romani, 2020 for further details).

⁵ Highlighted in bold are the metonymic semantic types that are also target types (for example, [Sound] is the metonymic semantic type in “[Sound] is emitted by [Human]”, but also the target semantic type in “[Medium] produces [Sound]”).

from ItWaC reduced corpus, for each relation found. For each instance, the metonymic argument (exemplifying the source-metonymic semantic type) is highlighted in bold, and the verb triggering the metonymy is in italics.

As a second step, we attempted to draw a map of the metonymic relations, by connecting the target semantic types to their metonymic arguments. In Figure 3, each target semantic type is at the centre of a circular area (target type area), highlighted in bold; in each area the metonymic types related to the target semantic type are included; for each target semantic type, a different colour is given. In most of the cases, they intersect with each other, showing how semantic types can refer to different areas. For example, [Sound] and [Human] share various semantic types (e.g. [Machine], [Musical Instrument], [Medium], as visible in Figure 3) as they can be used both to refer to [Sound] and to [Human] (for clarifying examples, see Table 2 in the Appendix). As mentioned, we included metonymic semantic types in the areas of the map. In our representation, metonymic and target semantic types are connected to each other through arrows, on which the relation is notated. The direction of the arrow traces the direction of the metonymic shift: from the metonymic semantic type to the target semantic type (e.g., [Container] → [Beverage]).

Our results show that metonymic semantic types are fluid; target types can also be metonymic types,

in certain contexts, as previously mentioned. For example, [Human] is the target type for [Document] (as [Document] is written or composed by [Human] as an author; e.g. ‘this book tells about II World War’) but also [Document] is the target type for [Human] (as [Human] writes or composes [Document]; e.g. ‘I am reading *Shakespeare*’).

The structure of the map we conceived draws attention to two main aspects. First, it depicts the complexity of the metonymic relations between semantic types and highlights how metonymy is not a unidirectional phenomenon but, conversely, it is fluid and changeable. Second, from a cognitive point of view, [Human] is at the centre of most of the relations and each target type area is connected to it by multiple relations. In particular, in our data, [Human] is deeply connected and involved within the [Sound] area (for more details, see Romani, 2020).

Finally, for what concerns the limited sample of verbs under investigation, it is interesting to notice that even if there are various source types, the potential target semantic types are only six. We may argue that there is a limited number of target types that attract different source types, in particular regarding [Human] and [Sound], which have the highest number of relations (see Table 2). Further investigation on this point is necessary, together with the extension of the number of examined verbs and instances.

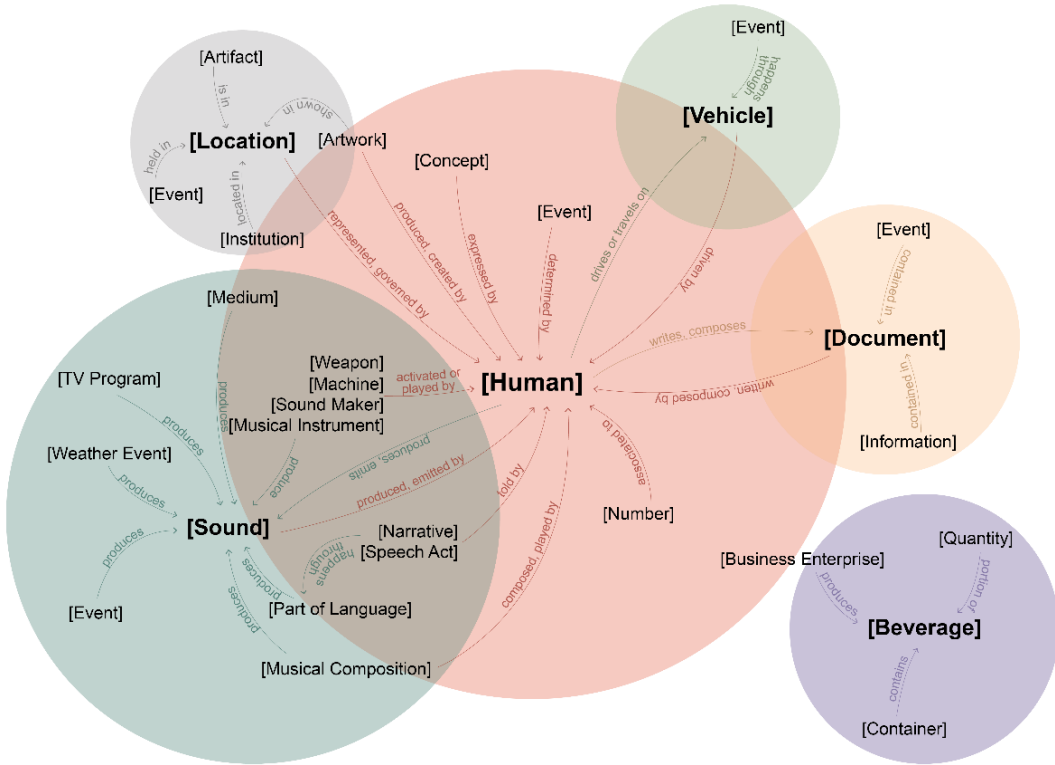


Figure 3. Map of metonymic relations between source and target semantic types, linked by an arrow

5 Conclusions and future works

In this paper, we approached the study of metonymy from a corpus-based perspective. The research was conducted on a selected list of verbs, taken from a background study (Ježek & Quochi, 2010). Our aim was to search for metonymic phenomena inside a corpus of Italian language and to register them in a resource for Italian verbs, T-PAS. To do so, we conceived an annotation scheme and procedure that gave us relevant results and allowed us to register a variety of metonymic relations.

We also attempted to make some theoretical generalizations based on the metonymic relations we found through the corpus analysis. We therefore created a list of metonymic relations and we designed a map in which the relations are connected to the semantic types they involve. Both the map and the list depict the complexity and variety of the phenomenon, in terms of number of possible metonymic relations and of the semantic types interested.

In future perspectives, we intend to enrich the map and the list with new relations by extending the number of verbs investigated and to evaluate the annotation procedure. For future annotations, we will provide the current version of the list and of the map on the online public version of T-PAS (upcoming). We are also interested in comparing our results with those in Marini & Ježek (2020), in a cross-linguistic perspective.

In line with previous studies (Section 2), we believe that the annotated corpus data, as well as the relations in Table 2, could be useful for automatic detection of metonymies. To our knowledge, little work has been done on this for Italian language: it would be therefore intriguing to test our data in NLP tasks.

References

- Baroni, M., Bernardini, S., Ferraresi, A. & Zanchetta, E. (2009). The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, 43(3):209–226.
- Fauconnier, G. (1985). *Mental Spaces*. Cambridge, USA: MIT Press.
- Hanks, P. (2013). *Lexical Analysis: Norms and Exploitations*. Cambridge: The MIT Press.
- Ježek, E. (2016). *The Lexicon: An Introduction*. Oxford: Oxford University Press.
- Ježek, E. (2019). Sweetening Ontologies Cont'd: Aligning bottom-up with top-down ontologies. In Barton, A., Seppälä, S. & Porello, D. (eds.), *Proceedings of the Joint Ontology Workshops 2019*. Graz, Austria.
- Ježek, E., Magnini, B., Feltracco, A., Bianchini, A. & Popescu, O. (2014). T-PAS: A resource of corpus-derived Types Predicate-Argument Structures for linguistic analysis and semantic processing. In *Proceedings of LREC*, 890-895.
- Ježek, E. & Frontini, F. (2010). From Pattern Dictionary to PatternBank. In de Schryver G.M. (ed.), *A Way with Words: Recent Advances in Lexical Theory and Analysis*. Menha Publishers, Kampala. 215-239.
- Ježek, E. & Quochi, V. (2010). Capturing Coercions in Texts: A First Annotation Exercise. In *Proceedings of the 7th International Conference on Language Resources and Evaluation. La Valletta, Malta*. (LREC-2010), 1464-1471.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., Rychlý, P. & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1(1), 7-36.
- Lakoff, G. & Johnson, M. (1980). *Metaphors we Live By*. Chicago, USA: University of Chicago Press.
- Marini, C. & Ježek, E. (2019). CROATPAS: A Resource of Corpus-derived Typed Predicate Argument Structures for Croatian. In *Proceedings of the 6th Italian Conference on Computational Linguistics (CLiC-it)*. Bari, Italy.
- Marini C. & Ježek, E. (2020). Annotating Croatian Semantic Type Coercions in CROATPAS. In *Proceedings of the 16th Joint ACL-ISO Workshop Interoperable Semantic Annotation (ISA-16)*, pages 49–58. *Language Resources and Evaluation Conference (LREC 2020)*, Marseille, May 2020.
- Markert, K. & Nissim, M. (2006). Metonymic Proper Names: A Corpus-Based Account. In Gries, S.T. & Stefanowitsch, A. (eds.), *Corpora in Cognitive Linguistics. Vol. 1: Metaphor and Metonymy*. Mouton-De Gruyter: Berlin & New York.
- Markert, K. & Nissim, M. (2007, June). Semeval-2007 task 08: Metonymy resolution at semeval-2007. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)* (pp. 36-41).
- McGregor, S., Ježek, E., Purver M. & Wiggings G. (2017). A Geometric Model for Detecting Semantic Coercion. In Gardent, C. and C. Retoré (eds) *Proceedings of the 12th International Workshop on Computational Semantics (IWCS)*, ACL Anthology, Association for Computational Linguistics, pp. 1-14.
- Pustejovsky, J. & Ježek, E. (2008). Semantic Coercion in Language: Beyond Distributional Analysis. In *Italian Journal of Linguistics/Rivista Italiana di Linguistica* 20(1).
- Pustejovsky, J., Rumshisky, A., Plotnick, A., Ježek, E., Batiukova, O. & Quochi, V. (2010). SemEval-2010

Task 7: Argument Selection and Coercion. In *Proceedings of the 5th International Workshop on Semantic Evaluation (ACL-2010)*. Uppsala, Sweden. 27-32.

Romani, E. (2020). *Searching for Metonymies in Natural Language Texts. A Corpus-based Study on a Resource for Italian Verbs*. BA Thesis, University of Pavia, Pavia, Italy.

Appendix

<i>metonymic (source) semantic type</i>	<i>relation</i>	<i>target semantic type</i>	<i>corpus example (ItWaC reduced)</i>
[Vehicle]	is driven by	[Human]	L'agente scese velocemente in strada, <i>chiamò un taxi</i> e dette l'indirizzo segreto.
[Document]	is written, composed by	[Human]	Il gioco, <i>informa un comunicato</i> , sarà lanciato contemporaneamente in Italia e Regno Unito.
[Location]	is represented, governed by	[Human]	Dissidenti e rifugiati <i>accusano</i> la Corea del Nord di tortura politica e [...] chiedono di includere nei colloqui anche il tema dei diritti umani e delle libertà fondamentali.
[Sound]	is produced, emitted by	[Human]	A questo punto <i>una voce interrompe</i> Gesù.
[Event]	is determined by	[Human]	Ricordo la telefonata che mi <i>raggiunse</i> la mattina presto nella mia abitazione di Milano, la corsa in ufficio, il viaggio dell'indomani nei luoghi della catastrofe [...]
[Projectile]	is shot by	[Human]	L' uomo viene <i>raggiunto</i> da cinque proiettili e muore mentre viene trasportato in ospedale.
[Sound Maker]	is activated by	[Human]	Una campana <i>annuncia</i> l'inizio della messa.
[Machine]	is activated by	[Human]	L'altoparlante <i>annunciava</i> l'arrivo di un treno.
[Musical Instrument]	is played by	[Human]	Oltre al Flauto d'oro, lo zufolo pastorale <i>annuncia</i> ed accompagna Papageno.
[Concept]	is expressed by	[Human]	Alcuni studiosi <i>accusavano</i> la psicologia di naturalismo, altri di non essere una scienza naturale.
[Number]	is associated to	[Human]	L'iniziativa consiste nella possibilità per gli anziani di <i>contattare un numero</i> messo a disposizione gratuitamente dal Comune, [...] che attiverà uno degli oltre mille volontari.
[Part of Language]	is pronounced by	[Human]	La frase venne interrotta dal suono di sirene, quelle della Polizia.
[Event]	is held in	[Location]	Una sera, mentre <i>si sta recando ad una cena</i> dove dovrà tenere un discorso, Henry riceve l'invito a presentarsi al commissariato.
[Institution]	is located in	[Location]	Giovanni Paolo II <i>ha visitato</i> il Parlamento italiano, su invito dei Presidenti della Camera dei Deputati.
[Artwork]	is shown in	[Location]	[...] raggiungiamo piazza Pio IX dove sorge la Pinacoteca Ambrosiana, entriamo per <i>visitare le opere</i> di Caravaggio, Leonardo e Botticelli.
[Artifact]	is in	[Location]	La mia peste la sento tre volte al giorno, anche se non vuole <i>venire al telefono</i> a parlarmi [...]
[Human]	writes, composes	[Document]	<i>Ho letto</i> Dante, Moravia, Calvino .
[Information]	is contained in	[Document]	Vi raccomandiamo, prima di procedere nella consultazione, di <i>leggere le avvertenze</i> .
[Event]	is contained in	[Document]	Consiglio di <i>leggere</i> senza paraocchi ideologici questa intervista del prof. Dallapiccola sulla diagnosi preimpianto.
[Container]	contains	[Beverage]	Al pub Orange Paolo <i>aveva bevuto un bicchiere</i> di troppo e alcuni clienti [...] hanno chiesto l'intervento dei carabinieri

			perché venisse allontanato.
[Quantity]	is a portion of	[Beverage]	Occorre portarsi le sedie e il fuoco, e mettere ciascuno due soldi, se si vuole <i>bere un goccio</i> .
[Business Enterprise]	produces	[Beverage]	Anche noi della Nazionale <i>beviamo Uliveto!</i>
[Human]	drives or travels on	[Vehicle]	Il presidente del Consiglio è <i>atterrato</i> a mezzogiorno sul campo sportivo di Sant'Agnello a Sorrento.
[Fantasy Character]	drives or travels on	[Vehicle]	Una coppia di alieni <i>atterra</i> sulla Terra, precisamente in una campagna.
[Event]	happens through	[Vehicle]	Alle 16:50 è <i>atterrato</i> il volo speciale Parigi-Beirut della linea di bandiera libanese.
[Machine]	activated by [Human], produces	[Sound]	Ma molti non hanno voluto <i>ascoltare</i> la sirena d' allarme e sono rimasti nelle loro abitazioni [...]
[Weapon]	activated by [Human], produces	[Sound]	I fulcili <i>echeggiano</i> in lontananza mentre tutto intorno continua a muoversi e girare.
[Sound Maker]	activated by [Human], produces	[Sound]	Le campane non risuoneranno i rintocchi della morte, ma <i>echeggeranno</i> a festa per celebrare la Vita.
[Musical Instrument]	played by [Human], produces	[Sound]	Le trombe non si <i>udivano</i> più, ma dalla parte della vallata si udivano ad intervalli dei lontani fragori.
[Human]	produces, emits	[Sound]	Se io fossi una persona che non ha mai <i>ascoltato</i> Patty Smith [...] magari mi passerebbe anche la voglia di andarla a scoprire.
[Weather Event]	produces	[Sound]	<i>Ascolta</i> la pioggia , se hai sonno ti tengo con me.
[Part of Language]	pronounced by [Human], produces	[Sound]	Avete <i>ascoltato</i> tutti le parole di Romano: sono sicuro che tanti tra noi pensano che le sue idee siano una buona base per governare il Paese.
[Narrative]	told by [Human] through [Part of Language], produces	[Sound]	Per <i>ascoltare</i> un racconto , una storia , occorre restare in silenzio.
[Speech Act]	told by [Human] through [Part of Language], produces	[Sound]	Low Key <i>udì</i> a stento la domanda di Eric mentre tornava a concentrarsi sul presente.
[Event]	involves	[Sound]	In una grotta dedicata alla Madonna di Lourdes è possibile, oltre che <i>ascoltare</i> la Santa Messa la domenica, celebrare matrimoni [...]
[Medium]	produces	[Sound]	Roberto Landi sta seduto dentro il camper e <i>ascolta</i> la televisione .
[TV Program]	emits	[Sound]	L'autista stava <i>ascoltando</i> un notiziario della Bbc su quanto è accaduto qualche giorno fa a Madaen.

Table 2. Metonymic relations (column 2) identified between the source (metonymic) semantic type (column 1) and the target semantic type (column 3), with an instance from ItWaC for each relation found (column 4)