

Valentina Bambini[#], Marta Ghio[♦], Petra B. Schumacher[°]

Acceptability differences between metaphor, metonymy and approximation*

(preliminary version of a paper submitted elsewhere for final publication)

Pragmatic and cognitive accounts of figurative language posit a difference between metaphor and metonymy in terms of underlying conceptual operations. Recently, other pragmatic uses of words have been accounted for in the Relevance Theory framework, among which approximation, described in terms of conceptual adjustment that varies in degree and direction with respect to metaphor. However, there is very few experimental evidence addressing the metaphor/metonymy distinction, and none concerning approximation. Here we used tightly normed materials to elucidate the interpretation mechanisms of these three phenomena through timed sensicality judgments. Results revealed that interpreting metaphor and approximation differs from literal interpretation both in accuracy and reaction times, with higher difficulty and costs for metaphor than for approximation. This seems to suggest similar albeit gradual interpretative costs, potentially linked to differences in the underlying conceptual adjustment processes, as hypothesized by Relevance Theory. Metonymy, on the contrary, exhibits a different style of interpretation in speeded conditions, which almost equates literal comprehension and calls for a theoretical distinction from metaphor. The idea of a routinization of the conceptual shift underlying metonymic

[#] Center for Neurolinguistics and Theoretical Syntax, Institute for Advanced Study, IUSS, Pavia, Italy & Laboratorio di Linguistica “G. Nencioni”, Scuola Normale Superiore, Pisa, Italy

[♦] Laboratorio di Linguistica “G. Nencioni”, Scuola Normale Superiore, Pisa, Italy

[°] Independent Emmy Noether-Research Group, Department of English and Linguistics, Johannes Gutenberg-University Mainz, Germany

* The current work is based on a collaborative research initiative funded by the European Science Foundation’s EURO-XPRAG Research Networking Program to the first and third author. The study was ideated and elaborated jointly by the three authors; data preparation, collection and analysis were conducted by the first and second author. We would like to thank Ira Noveck and Pier Marco Bertinetto for invaluable discussion and Chiara Bertini for assistance during data collection. Thanks goes also to Lewis Bott for helpful suggestions aimed at an ongoing collaborative development of the project.

uses is also discussed. Overall, this work represents a first attempt to provide an empirical basis for a theory-sound and psychologically-grounded taxonomy of figurative and loose uses of language.

1 Introduction

Word meaning is often modified in use, giving rise to a number of loose and figurative uses. These modulations of meaning are thoroughly context-dependent and as such their description has fallen under the domain of pragmatics. While it is useful to group non-literal uses under the same family, potential differences in representation and the underlying interpretative mechanisms should be accounted for and considered experimentally. This paper is concerned with the characterization of three pragmatic phenomena, namely metaphor (e.g., ‘Some theses are marathons’), metonymy (e.g., ‘No comments from Buckingham Palace’) and approximation (e.g., ‘The room is empty’), and whether they exhibit different interpretation costs, which might support and sharpen theoretical distinctions.

Existing pragmatic and cognitive accounts differ on whether they treat different types of pragmatically enriched meanings as distinct operations or not. In the Gricean framework, figurative expressions such as metaphor, irony, meiosis and hyperbole are grouped together as cases of flouting the first Maxim of Quality, i.e., “Do not say what you believe to be false”, which require the derivation of an implicature (Grice 1975). The nature of metonymy is not explored, but it can reasonably be assumed that also metonymic expressions would be described inferentially, either as another case of flouting the first Maxim of Quality or as a tool to adhere to the Maxim of Manner (Egg 2004). In more recent times, Relevance Theory has deepened the consideration of the inferential processes underlying the comprehension of the lexical items, suggesting that grasping the intended meaning of a word requires a process of adjusting the linguistically encoded concept to construct an ad hoc concept (notated with an asterisk), i.e., a concept inferentially derived for that occasion of use, whose denotation is broader (i.e., more inclusive) or narrower (i.e., less inclusive) than the denotation of the lexical concept. Consider the utterance ‘Boris is a man’: in most contexts the lexically encoded concept MAN would result underinformative and require, for instance, to be narrowed down to MAN* as ‘ideal man’, in order to reach the intended interpretation. Conversely, in ‘This policy will bankrupt the farmers’, the encoded concept BANKRUPT could be taken literally, but in certain contexts is likely

to require an adjustment that goes in the opposite direction, namely to be broadened in order to include cases in which the farmers are close enough to bankruptcy (Carston 2009). Importantly, both narrowing and broadening are the outcome of an adjustment process that meshes content, context, and cognitive effects, constrained by expectations of relevance raised by the utterance itself (Wilson 2003: 356). For instance, the same sentence ‘Boris is a man’, uttered in a context in which Boris is watching a football match, would produce a narrowing of the denotation of MAN to MAN* as ‘typical man’.

In this view, most words require an adjustment process resulting in ad hoc concept construction, and the difference depends on the direction of the adjustment (broadening or narrowing with respect to the denotation, as shown above), and also on the degree of it, ranging from cases where the ad hoc concept is present in the hearer’s cognitive system to cases that are highly context-dependent and entirely occasion-specific (Wilson 2003; Sperber & Wilson 2005; Wilson & Carston 2006). Well studied examples of different degrees of broadening include language uses known in the literature with the labels of approximation, hyperbole and metaphor. Approximation is a variety of broadening that includes a relatively marginal adjustment of the encoded concept, to cover just a ‘penumbra’ of cases that only strictly speaking fall outside the linguistically-specified denotation. Hyperbole involves a more substantial adjustment of the encoded concepts. And metaphor is a use of language based on an even more radical broadening of the lexical concept. The following example might be especially helpful for illustrating the gradient of conceptual broadening: consider the utterance ‘The house is empty’. In some contexts the encoded concept EMPTY might be used approximately, to communicate the slightly weaker claim that the house is lacking of furniture. In other circumstances, the same encoded concept EMPTY might be intended more weakly, as claiming that the house, although furnished, doesn’t have as much furniture as desired, resulting in a case of hyperbole. Depending on context, it is also possible that the intended meaning makes a weaker claim still, more distant from the lexical concept, namely a metaphor, communicating for instance that the house lacks of emotional content (for similar examples, Wilson & Carston 2007; Carston 2010).

Recently, Carston & Wearing (2011) proposed a finessing of the relevance-theoretic account, by positing a stronger distinction between metaphor and hyperbole: while concept broadening is required both in metaphor and hyperbole understanding, metaphorical uses would require also additional concept narrowing. Consider the

following utterances: ‘My evening jog with Bill turned into a marathon’ and ‘Writing a thesis was a marathon Jane didn’t want to repeat’. When intended hyperbolically, the denotation of the ad hoc concept MARATHON* is simply more inclusive (broader) than that of the original lexical concept, involving a relaxing of the length of the episode of running. When intended metaphorically, the word goes through a broadening (in order to include instances of activities that are psychologically demanding and exhausting) combined with narrowing (in order to exclude professional marathons) (Carston & Wearing 2011: 286-296).

Following this idea, it seems reasonable to assume that also approximation differs from metaphor in requiring only (and marginally) concept broadening: if a separation holds between the case of metaphor and the case of a substantial yet not radical broadening such as hyperbole, the separation should hold also between metaphor and a marginal broadening like that required by approximation. A sentence such: “The Foundation organizes a marathon for children”, where MARATHON* is only marginal more inclusive than the lexical concept, indicating that the children won’t run for 42 km but most else equals an athletic marathon, should count as an example for this. Along this line, approximation and metaphor would differ not only in degree but also in direction of the adjustment process.

As concerns metonymy, a full description in Relevance Theory terms is still lacking (but see Papafragou 1996 for a preliminary account), and indeed considered as an interesting challenge for pragmatics (Carston 2010). The intuition put forward up to now is that metonymy, while still representing a pragmatic use that requires the construction of an ad hoc concept, does not involved the same kind of conceptual adjustment observed for metaphor. In the case of metonymies like ‘The ham sandwich wants to pay’, the concept HAM SANDWICH cannot be broadened nor narrowed to include the person who ordered a ham sandwich. Rather, the procedure is likely to include some kind of real world ‘association’ or genuine reference substitution, not straightforwardly reducible to lexical narrowing or broadening (Wilson & Carston 2007; Carston 2010: p. 160). As such, metonymy cannot be accounted for in the frame of degree and direction of broadening / narrowing, but would rather call for a different interpretative style, based on different interpretative patterns, presumably of a meaning shift type. The idea of a transfer of meaning is in line with other pragmatic

distinguishing between loosening, meaning shift, and free enrichment (Recanati 2010), and also with the standard definition of metonymy (Nunberg 1995)¹.

Turning to the framework broadly known as Cognitive Linguistics, both metaphor and metonymy are thought of as conceptual phenomena grounded in general cognition. Yet a difference is assumed to characterize the two: while metaphor is described in terms of mapping between two distinct cognitive domains (e.g. in ‘Love is a journey’, the source domain JOURNEY is mapped onto the target domain LOVE), metonymy is based on mapping within the same cognitive domain (e.g., in ‘He is reading Shakespeare’, the source domain SHAKESPEARE provides access to its sub-domain SHAKESPEARE’S WRITINGS, which is the target domain) (Lakoff & Johnson 1980; Ruiz de Mendoza Ibáñez 2007). Furthermore, the mapping is taken to be based on different associative relations: resemblance for metaphor and contiguity for metonymy. Standard types of metonymic mappings are, among others, part for whole, producer for product, place for institution, object used for user (Panther & Thornburg 2007). This list suggests the routinized status of many metonymic mappings, and points in the direction of a close relation between metonymy and grammar. There are indeed grammatical structures that seem to be sensitive to metonymically induced interpretations. For example, in ‘The author began the book’, the verb’s logical structure coerces an interpretation in which a part of an event, the book, denotes the whole event, writing the book, a phenomenon known as “logical metonymy” (Pustejovsky 1995; Lascarides & Copestake 1998).

Do the differences among pragmatic uses brought about in the theoretical literature find experimental support? Evidence from direct comparison of metaphor and metonymy is sparse. In a self-paced reading study, Gibbs (1990) showed that metaphorical referential descriptions are understood more easily than metonymic ones. Developmental psychology, however, points in a different direction: with respect to metaphor, metonymy is acquired earlier and processed more accurately, and possibly associated to a more basic conceptual and cognitive pattern (Rundblad & Annaz 2010). Also, metonymy comprehension correlates with the expansion of receptive

¹ Following Nunberg (1995)’s distinction between reference transfer (‘The ham sandwich wants to pay’, where an NP is used to refer to another NP) and predicate transfer (‘Nixon bombed Hanoi’, where the shift concerns the whole predicate ‘bombed Hanoi’), Wilson & Carston (2007) suggested that only the former case involves genuine substitution, while the latter should be accommodated in the ad hoc concept perspective, through broadening and narrowing. Since the focus of our paper is the reference transfer case, this interesting suggestion will not be further discussed here.

vocabulary, which suggests that, at least for some cases, metonymic expressions might be part of the lexical knowledge (Annaz et al. 2009).

In considering how different pragmatic uses are processed, we can nevertheless rely on the extensive literature on metaphor and – to a lesser extent – metonymy. It has been shown that metaphor processing is influenced by many factors, such as context, familiarity, difficulty, novelty (Gibbs 1994; Giora 2003; Cardillo et al. 2010). However, when placed in a minimal context and controlled for the other factors, processing metaphorical expressions still requires additional effort measured both at the behavioral level (Cacciari & Glucksberg 1994; Noveck et al. 2001) and in terms of brain response (De Grauwe et al. 2010; Bambini et al. 2011). Metonymy, on the contrary, has produced mixed results. Neurophysiological and neuroimaging evidence for a difference between metonymy and literal processing (Schumacher 2011; Rapp et al. 2011) stands aside with eye-tracking studies showing no differences with literal interpretations when metonymies are licensed by the context and mediated by a common metonymic convention (Frisson & Pickering 1999; Frisson 2009). It is also noteworthy that the investigations at the neural level mentioned above failed in finding significant differences in terms of behavioral response (Schumacher 2011; Rapp et al., 2011). Up to now, to the best of our knowledge, approximation hasn't received empirical consideration. However, it has been shown that other types of loose use, such as hyperbole, are read faster than metaphor (Deamer et al. 2010), lending support to the gradient of meaning extension.

In the perspective of experimental pragmatics (Noveck & Reboul 2008), the purpose of this study is to compare interpretation availability and costs for different types of pragmatically modulated meaning, i.e., metaphor, metonymy, and approximation, in order to provide empirical evidence in favor of either a distinction or a unified view of the three phenomena. In a rating study, we de novo built a set of metaphors, metonymies and approximation (with corresponding literal and anomalous counterparts), which were further rated for the major psycholinguistic properties, namely meaningfulness, difficulty, cloze probability and familiarity, in order to obtain a well-balanced pool of stimuli especially controlled for their interpretability. Normed stimuli for the three pragmatic phenomena were then tested through a timed sensibility judgment paradigm, where participants are asked to decide quickly if a sentence is meaningful or not, and their performance is measured in terms of accuracy and reaction times. This experimental paradigm seems especially suitable to explore pragmatic modulation, as it requires subjects not only to access the linguistic items but

to elaborate and interpret their meanings at the level of detail that would distinguish different senses (Klein & Murphy 2001). Following Relevance Theory, metaphor and approximation are expected to show a gradient of availability and costs in comparison with literal interpretation, as a result of different degrees and directions in the underlying conceptual adjustment processes. Specifically, the combination of broadening and narrowing for metaphor should reflect in higher interpretative costs with respect to the marginal broadening required by approximation. Metonymy, on the contrary, should exhibit a dissimilar style of interpretation, as can reasonably be hypothesized on the basis of the specific mechanisms posited by Relevance Theory and Truth-conditional Pragmatics, not reducible to broadening and narrowing of the lexical concept but rather based on a sort of meaning shift. This second hypothesis is indeed compatible with the Cognitive Linguistics account too, where a within vs. between domains mapping distinction is drawn. Below we will first present detailed background on the construction of the stimulus material and the rating study, and then go back to the kinds of conceptual adjustments and how they might result in different sensicality judgment responses.

2 Rating study

In building an experimental set of different pragmatic uses, two major issues emerge: first the need to rule out possible confounding effects due to sentential and contextual environment, and second the need to control for a number of psycholinguistic variables that are well known for influencing figurative language.

As for the first point, for each of the three pragmatic phenomenon under investigation (metaphor, metonymy and approximation) we constructed a set of Italian sentences of the form ‘That Y verb X’, where X is the word triggering the pragmatic interpretation and is taken as the target word for the experimental measures. Literal and anomalous counterparts were created for each set by selecting different subject nouns ‘Y’ or different verbs (see below for set specific criteria). Given the well-documented role of context in facilitating figurative language processing (Gibbs 1994; Schumacher 2012), across sets context was set to a minimal yet sufficient level for interpretation, in order to allow distinct pragmatic mechanisms to emerge neatly.

Among the many variables involved in figurative language, the three sets were rated for meaningfulness, difficulty, cloze probability and familiarity. The importance

of these variables has been extensively described for metaphor processing (Kintsch & Bowles 2002; Cardillo et al. 2010), partially addressed for metonymy processing (Frisson & Pickering 1999), and never explored for approximation. In the perspective of the timed sensicality judgments to be collected afterwards, the main purpose of this rating study was to assess the meaningfulness of the experimental items, i.e., the interpretability of the sense of the utterances, along with their difficulty, i.e., the overall ease of interpretation. Accordingly, our goal was twofold. On the one hand, for each set we aimed at having pragmatic uses as interpretable as literal uses when participants are given as much time as they want, and we aimed at separating them from anomalies, in order to exclude the possibility that potential pragmatic effects arising in the timed sensicality judgment could be due to meaningfulness variance. On the other hand, we aimed at having three sets equally characterized for meaningfulness, in order to guarantee comparability across phenomena. Overall, it was important to maintain a low level of difficulty in assigning meaning.

As for familiarity, we aimed at setting a medium level of familiarity for the pragmatic uses, in order to exclude both fully conventionalized expressions that could be processed as idioms rather than through pragmatic adjustment and highly creative expressions that could demand special pragmatic processes or even result senseless. The familiarity dimension was differently operationalized for each pragmatic phenomenon under investigation. Following the main literature on metaphor, familiarity was assessed by asking participants to rate frequency of experience for each metaphorical sentence. For the metonymy set, we devised a world-knowledge task to control for both the familiarity of the names used as target words and the familiarity of the metonymic transfer for those names. For the approximation set, a typicality task was used, where participants rated how appropriately the adjectives used as target words X qualifies the subject words Y.

Lastly, to ensure that context was kept minimal and equal across conditions, we tested the contextual expectancy of each target word X for each sentence in the triplet and for all sets through a cloze probability task.

2.1 Methods

Participants

Eighty-five native speakers of Italian (42 F /43 M, mean age = 26.85 ± 3.80 , mean schooling years = 18.02 ± 2.04 years of education) completed the

questionnaire. Participants were unaware of the aim of the questionnaire and were not informed about the inclusion of figurative language. They gave written consent to participate after receiving an explanation of the procedures, according to the Declaration of Helsinki.

Materials

For each phenomenon under consideration (metaphor, metonymy and approximation) we constructed a set of forty-eight triplets including sentences with the pragmatic use, literal and anomalous counterparts (henceforth, for the sake of brevity, the label ‘pragmatic sentences’ will be used to refer to the pragmatic use condition for each set). The triplets were designed according to the criteria below, resulting in a total pool of 432 sentences. Table 1 shows an example of triplets from each set (Metaphor set, Metonymy set, Approximation set).

	Pragmatic	Literal	Anomalous
Metaphor set	Quelle ballerine sono farfalle	Quegli insetti sono farfalle	Quelle bottiglie sono farfalle
	<i>Those dancers are butterflies</i>	<i>Those insects are butterflies</i>	<i>Those bottles are butterflies</i>
Metonymy set	Quello studente legge Camilleri	Quel giornalista intervista Camilleri	Quel cuoco cucina Camilleri
	<i>That student reads Camilleri</i>	<i>That reporter interviews Camilleri</i>	<i>That chef cooks Camilleri</i>
Approximation set	Quelle gomme sono lisce	Quel marmo è liscio	Quei ristoranti sono lisci
	<i>Those tires are smooth</i>	<i>That marble is smooth</i>	<i>Those restaurants are smooth</i>

Table 1. Examples of stimulus triplets for the Metaphor set, the Metonymy set and the Approximation set. Original Italian; English translation in italics.

Metaphor set. We constructed nominal metaphors where a noun X is the vehicle for the metaphorical meaning (e.g., ‘Those dancers are butterflies’). For each noun, one literal sentence (e.g., ‘Those insects are butterflies’) and one anomalous sentence (e.g., ‘Those bottles are butterflies’) was created. The copular construction was chosen in order to reduce the effects of verb semantics. Plural forms were used in order to avoid

predictability effects carried by the gender-marked articles required in the singular forms.

Metonymy set. We built a set of producer-for-product metonymies, where proper names of well-known Italian people were metonymically used to refer to objects. Different types of producer for product shift were used, such as author for book (e.g., ‘That student reads Camilleri’), musician for song, designer for manufacture, painter for painting. In terms of Ruiz de Mendoza Ibáñez’s taxonomy, all metonymies were of the type target-in-source, i.e., the product is a subdomain of the producer (Ruiz de Mendoza Ibáñez 2007). Each proper name X was also combined with different subject nouns Ys and different verbs, resulting once in a literal sentence (e.g., ‘That journalist interviews Camilleri’) and once in an anomalous sentence (e.g., ‘That chef cooks Camilleri’). In order to confine metonymic interpretation to the target word, subject nouns and verbs were syntactically and semantically congruent in all conditions. Names of presently popular Italian people (e.g., Camilleri, Saviano, Vasco) were chosen instead of very famous people from the past (e.g., Dante, Verdi) in order to reduce conventionality, as it has been shown that the use of famous names (e.g., Dickens) in the metonymic form has become lexicalized in ordinary language (Frisson & Pickering 2007).

Approximation set. Among the different cases of approximate uses (Wilson 2003; Wilson & Carston 2006), we focused on adjectives. Following the examples provided by Wilson & Carston (2007), four main types of adjectives used in an approximate fashion were included: sense-related (e.g., ‘Those tires are smooth’), geometric-related (e.g., ‘Those sunglasses are rectangular’), color-related (e.g., ‘Those clouds are black’), and negative-related adjectives (e.g., ‘Those strawberries are tasteless’). For each target word X, we created a literal sentence by selecting a prototypical exemplar having the property described by the adjective (e.g., ‘Those marbles are smooth’), and an anomalous sentence (e.g., ‘Those restaurants are smooth’). As in the metaphor set, all sentences were copular constructions in order to reduce the influence of verb semantics.

Tasks

Meaningfulness and Difficulty tasks. We asked participants to rate on a five-point Likert scale how meaningful each sentence was (1 = meaningless; 5 = very meaningful). Each sentence was presented one at a time, and participants selected the

value of the scale representing their judgment. Next, participants were asked to rate how difficult it was to rate the meaningful for that item, on a scale from 1 (very easy) to 5 (very difficult). All sets were tested.

Familiarity task. For each item in the Metaphor set, participants were instructed to indicate the frequency of experience with the sentence on a Likert scale from 1 (very unfamiliar) to 5 (very familiar). All sentences in the metaphorical set were tested.

World Knowledge task. For each proper name used in the metonymy set, participants were instructed to associate the proper name with the corresponding product, choosing between four options. The options vary according to the type of metonymy (e.g., for Camilleri, the options were: book / song / movie / painting). This should account for both the familiarity of the proper names and the familiarity of the producer-for-product transfer.

Typicality judgments task. We asked participants to indicate how appropriate a given adjective (e.g., smooth) is to qualify three different nouns (e.g., marble, tires and restaurants), which corresponded to the nouns used in the Approximation set. A 5-point Likert scale (1 = very inappropriate; 5 = very appropriate) was available for each noun. This should assess both the familiarity of the approximate use and the literal use. All triplets in the Approximation set were tested.

Cloze probability task. Each sentence was truncated before the target word, and participants were asked to complete with the first word that came to mind. All sets were tested.

Procedure

To preserve a high level of attention and avoid fatigue, two different questionnaires were created. Questionnaire 1 included three tasks: meaningfulness coupled with difficulty, world knowledge and typicality. Questionnaire 2 included cloze probability and familiarity tasks. For each questionnaire, the pool of 432 sentences was inserted into six different lists. Number of pragmatic, literal and anomalous sentences from each set was equally subdivided in the different lists and tasks. The lists were rotated among tasks so that each sentence was judged only once by each participant. The order of the tasks was counterbalanced across participants using a Latin Square procedure. Within each task the order of sentences was randomized. Half of the

participants completed one of the six lists of Questionnaire 1, the other half completed one of the six lists of Questionnaire 2 (number of data points per item per task > 6).

Ratings were administered online through Survey Monkey software (SurveyMonkey.com, LCC, Palo Alto, California, USA, www.surveymonkey.com). Each participant completed the questionnaire individually on a computer console, after giving informed consent through the same on-line procedure and reading online instructions. Each questionnaire lasted approximately 30 minutes.

2.2 Results

Inclusion criteria

Since the main aim of the ratings was to ensure the interpretability of the pragmatic sentences in untimed conditions for the purpose of the timed sensicality judgment task, we excluded pragmatic sentences with both median score equal to 1 on the meaningfulness scale and median score ≥ 3 on the difficulty scale. Set-specific criteria were also adopted in combination: metaphors that were judged very familiar (Mdn = 5) or very unfamiliar (Mdn = 1) were excluded; for metonymy, cases in which less than 80% of participants correctly associated the producer with the corresponding product were excluded; approximations for which the adjective-nouns pair scored < 2 (Mdn) on the typicality judgment scale were excluded. Literal and anomalous counterparts of the excluded pragmatic sentences were dropped as well.

From the original pool of sentences, 6 triplets were eliminated from each set. Final stimuli comprised 42 triplets (pragmatic, literal, anomalous sentences) for each of the 3 sets (metaphor, metonymy, approximation), resulting in a total of 378 sentences. In the following, we only report rating results for the final pool of sentences, to be further employed in the timed sensicality task.

Linguistic measures

Since the target word X was constant in the pragmatic, literal, and anomalous sentences of each triple, length and frequency were exactly balanced within each set. Length of the target words was also balanced across sets (metaphor: mean number of characters = 7.07; metonymy: mean number of characters = 7.08; approximation: mean number of characters = 7.47; $F(2,123) = 1.72$, $p = 0.18$). Frequency of the target words were controlled for metaphor and approximation based on the CoLFIS

corpus and frequency dictionary of written Italian (Bertinetto et al. 2005) (metaphor: mean log frequency = 1.50; approximation: mean log frequency = 1.54; $F(1,82) = 0.54$, $p = 0.81$). No values were available in the database for the proper names used in the metonymy sets. Ratings collected in the world knowledge task should suffice as a measure of subjective frequency (see below). Overall, average frequency based on the values of all content words of the sentence was balanced across sets (metaphor: mean log frequency = 2.33; metonymy: mean log frequency = 2.17; approximation: mean log frequency = 2.22; $F(2,123) = 0.75$, $p = 0.47$).

Rating results

We applied nonparametric methods since the assumptions underlying the use of parametric tests were violated in our sets². For descriptive statistics, we used median as a measure of central tendency and interquartile range as a measure of dispersion. For each set, Kruskal-Wallis test was performed to assess whether there were overall differences across pragmatic, literal and anomalous sentences (pragmatic modulation factor). We used post hoc Mann-Whitney U tests with Bonferroni correction for multiple comparisons (true alpha level = 0.0167) to determine which of the three types of sentences differed from each other. Meaningfulness and difficulty data were further analyzed across sets through correspondence analysis (PASW Statistics 18.0.0), an explorative computational method for interpreting categorical variables (Greenacre 1993). Data are organized into a contingency table that is visualized in the form of a spatial map of points representing the rows and the columns of the table: row and column points that are close together in the map are more alike than points that are far apart. Specifically, to evaluate the overall similarity/dissimilarity among the different types of sentences across sets, ratings were tallied in the form of a contingency table with the Likert scale scores in columns and all the sentence types in rows. Row principal normalization was applied; consequently, Euclidean distances in the spatial map between row points equal the chi-square distances between rows of the table.

² Whether Likert data should be analyzed with a parametric statistics ('liberal' approach) or nonparametric statistics ('conservative' approach) is matter of debate (Knapp 1990; Jamieson 2004; Carifio & Perla 2008; Norman 2010). We opted for the non-parametric version. However, we also conducted parametric statistics on rank transformed data (Conover & Iman 1981), by applying Univariate General Linear Model. In all cases, the results confirmed those obtained with the nonparametric procedure, and will not be reported in the results section.

Table 2 presents descriptive statistics of the key psycholinguistic variables computed for each set.

		Pragmatic	Literal	Anomalous
Metaphor set	<i>Meaningfulness</i>	4 (2-4)	4 (4-5)	1 (1-2)
	<i>Difficulty</i>	2 (1-2)	1 (1-2)	1 (1-2)
	<i>Familiarity</i>	3 (1-4)	4 (3-5)	1 (1-1)
	<i>Cloze Probability</i>	0.00%	0.39%	0.00%
Metonymy set	<i>Meaningfulness</i>	4 (3-5)	5 (4-5)	1 (1-2)
	<i>Difficulty</i>	1 (1-2)	1 (1-2)	1 (1-2)
	<i>World Knowledge</i>	90.29%	-	-
	<i>Cloze Probability</i>	0.00%	0.00%	0.00%
Approximation set	<i>Meaningfulness</i>	4 (3-5)	5 (4-5)	2 (1-2)
	<i>Difficulty</i>	1 (1-2)	1 (1-1)	2 (1-2)
	<i>Typicality</i>	4 (3-4)	5 (4-5)	1 (1-1)
	<i>Cloze Probability</i>	1.66%	11.44%	0.39%

Table 2. Descriptive statistics of rating scores for the Metaphor set, the Metonymy set and the Approximation set. Median and interquartile range (in brackets) are reported for meaningfulness, difficulty, and typicality tasks. Cloze probability and world knowledge results are reported in percentage.

Metaphor set

Both metaphorical and literal sentences scored median 4 on the meaningfulness scale (metaphorical: Mdn = 4, iqr = 2-4; literal: Mdn = 4, iqr = 4-5), while anomalous sentences were rated as meaningless (Mdn = 1, iqr = 1-2). The effect of pragmatic modulation was found significant ($\chi^2(2) = 432.84$, $p < 0.001$). Metaphorical and literal sentences significantly differed from anomalous counterparts (metaphorical vs. anomalous, $p < 0.001$; literal vs. anomalous, $p < 0.001$). Although both metaphorical and literal scores were in the upper end of the scale, statistically literal sentences resulted more meaningful than metaphorical sentences (literal vs. metaphorical, $p < 0.001$), probably due to a greater dispersion for metaphors. In all cases, participants formulated their judgments about the sense/nonsense of the sentences with no difficulty (metaphorical: Mdn = 2, iqr = 1-2; literal: Mdn = 1, iqr = 1-2; anomalous: Mdn = 1, iqr = 1-2), although there was an effect of pragmatic modulation ($\chi^2(2) = 29.59$, $p < 0.001$), due to higher scores for metaphor (metaphorical vs. literal, $p < 0.001$; metaphor vs. anomalous, $p < 0.001$). Familiarity

ratings showed that metaphorical sentences received medium values (Mdn = 3, iqr = 1-4). Literal sentences scored higher (Mdn = 4, iqr = 3-5), and the difference was significant ($\chi^2(2) = 495.71$, $p < 0.001$; literal vs. metaphorical, $p < 0.001$). Cloze probability was very low throughout the set, scoring 0.00% for metaphorical and anomalous sentences, and 0.39% for literal sentences.

Metonymy set

Metonymic and literal sentences received high scores on the meaningfulness scale (metonymic: Mdn = 4, iqr = 3-5; literal: Mdn = 5, iqr = 4-5), while anomalous sentences scored median 1 (iqr = 1-2). Pragmatic modulation yielded a significant effect ($\chi^2(2) = 446.02$, $p < 0.001$), with metonymic and literal items more meaningful than anomalies (metonymic vs. anomalous, $p < 0.001$; literal vs. anomalous, $p < 0.001$). As in the metaphor set, scores for both the metonymic and the literal items were at the upper end of the scale, but the comparison was significant (literal vs. metonymy, $p < 0.001$). Difficulty was very low across conditions (in all cases, Mdn = 1, iqr = 1-2). Nevertheless, we observed an effect of pragmatic modulation ($\chi^2(2) = 14.69$, $p < 0.001$), and the comparison between metonymies and literal sentences was significant ($p < 0.001$). World knowledge task showed that participants correctly associated the producer with the product (accuracy = 90.29%). Cloze probability was 0.00% for any version of any item.

Approximation set

On the meaningfulness scale, both approximations and literal sentences received high scores, while anomalous sentences received low scores (approximate: Mdn = 4, iqr = 3-5; literal: Mdn = 5, iqr = 4-5; anomalous: Mdn = 2, iqr = 1-2). A significant effect of pragmatic modulation was found ($\chi^2(2) = 451.60$, $p < 0.001$): both approximation and literal sentences were judged more meaningful than anomalies (approximate vs. anomalous, $p < 0.001$; literal vs. anomalous, $p < 0.001$). Consistently with findings on metaphor and metonymy, literal sentences were more meaningful than approximation sentences ($p < 0.001$). Difficulty was low throughout the set (approximate: Mdn = 1, iqr = 1-2; literal: Mdn = 1, iqr = 1-1; anomalous: Mdn = 2, iqr = 1-2), although the comparison between approximation and literal sentences reached significance ($\chi^2(2) = 101.89$, $p < 0.001$). Results of the typicality task showed that the adjectives were judged moderately appropriate when referred to

the nouns used in the approximations (Mdn = 4, iqr = 3-4), and fully appropriate when referred to the nouns used in the literal sentences (Mdn = 5, iqr = 4-5), while they were rated inappropriate in combination with the nouns from the anomalous sentences (Mdn = 1, iqr = 1-1). All comparisons were significant ($\chi^2(2) = 560.61$, $p < 0.001$; p 's < 0.001). Cloze probability remained below the threshold of 12%, with averaged values of 1.66% for approximations, 11.44% for literal and 0.39% for anomalous expressions.

Meaningfulness and difficulty across sets

A synthetic view of the similarity among different types of sentences with respect to meaningfulness and difficulty is provided through the correspondence analysis. For meaningfulness, a significant model was generated ($\chi^2(2) = 1762.36$, $p < 0.001$). The first two dimensions accounted for 84.9% and 11.6% of the total inertia respectively, as shown in Figure 1 (left panel). The spatial map revealed a clear segregation of the different types of sentences into two clusters on the first dimension: literal and pragmatic sentences are on the right side of the map, whereas anomalous sentences are on the left side. Hence, Dimension 1 seems to reflect meaningfulness, indicating that both pragmatic and literal sentences are similarly meaningful and differ from anomalous sentences. The second dimension seems to separate literal and pragmatic sentences. Dimension 2 could thus reflect the pragmatic modulation, suggesting a dissimilarity between the group of literal sentences and the group of pragmatic sentences. There is also an evident similarity among metaphorical, metonymic, and approximate sentences on the one hand, and all the literal counterparts on the other hand. The correspondence analysis was also applied to difficulty ratings. A significant model was generated ($\chi^2(2) = 277.89$, $p < 0.001$). The first two dimensions explain 61.7% and 32.0% of the total inertia, respectively. As shown in Figure 1 (right panel), all row points are close together, suggesting that all sentence types were perceived as similar for difficulty. In this case, the two dimensions are more difficult to interpret because they do not correspond to a clear pragmatic modulation, while the overall easy of interpretation of all sentence types seems to be clearly represented.

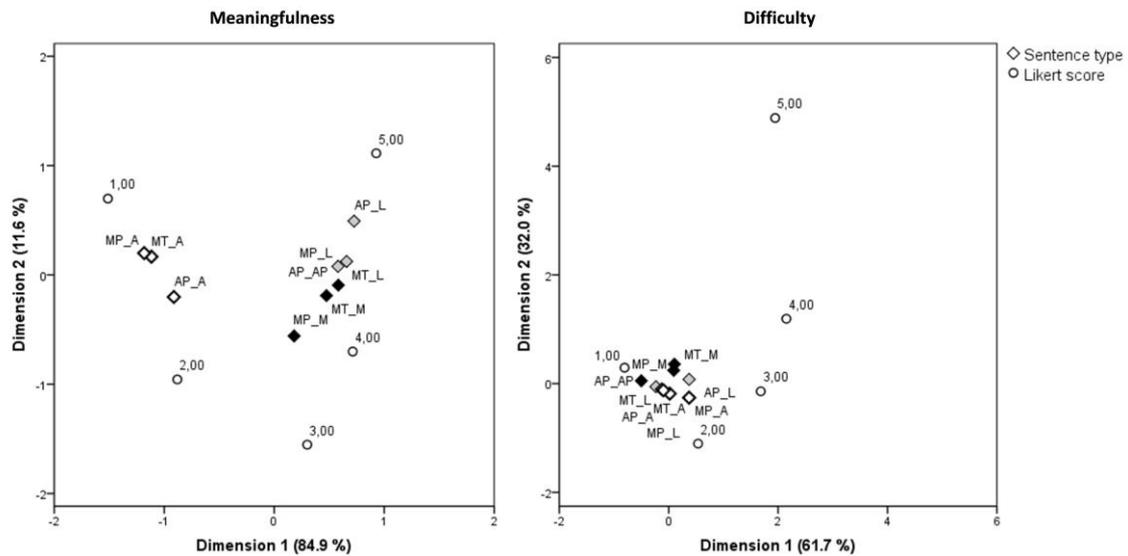


Figure 1. Two-dimensional correspondence analysis solution for Meaningfulness (left panel) and Difficulty (right panel) data. *MP_M*: Metaphor set – metaphorical sentences; *MP_L*: Metaphor set – literal sentences; *MP_A*: Metaphor set – anomalous sentences; *MT_M*: Metonymy set – metonymic sentences; *MT_L*: Metonymy set – literal sentences; *MT_A*: Metonymy set – anomalous sentences; *AP_AP*: Approximation set – approximate sentences; *AP_L*: Approximation set – literal sentences; *AP_A*: Approximation set – anomalous sentences. Pragmatic sentences are represented by black diamonds, literal sentences by grey diamonds, and anomalous sentences by white diamonds.

2.3 Discussion

Through a rating procedure, we built a set of tightly controlled metaphorical, metonymic and approximate uses (and literal and anomalous counterparts) based on the same sentence structure ‘That X verb Y’, controlled in terms of meaningfulness and difficulty. Inferential analysis revealed that both pragmatic and literal sentences were rated as meaningful and differently from anomalous sentences, thus excluding the possibility that pragmatic items are interpreted as anomalous. Inferential analysis also pointed out differences within the group of meaningful sentences, with literal uses scoring higher in meaningfulness and lower in difficulty than pragmatic uses. This is probably the price to pay for having set a minimal contextual environment that prompts pragmatic mechanisms to emerge. However, these differences do not jeopardize the general consistency of the sets. Accordingly, the correspondence analysis on meaningfulness shows that literal and pragmatic sentences of the three sets

clustered together and were judged similarly for meaningful, clearly differing from anomalous sentences. It also showed the similarity among all pragmatic uses, as opposed to all literal counterparts. Furthermore, all sentences were judged similarly on the difficulty scale, suggesting that our stimuli were all easily interpretable. Overall, the result of the ratings allows us to assume that potential differences in processing pragmatic and literal sentences in the timed sensicality judgments will not depend on differences in interpretability and difficulty of materials, but will truly reflect distinct interpretations for the three phenomena.

All pragmatic uses were also controlled for familiarity. Metaphorical sentences received medium scores on the familiarity scale, which suggests that they were not perceived either as fully conventionalized nor as extremely creative. Likewise, for metonymy, participants correctly associated the proper names of the producers to the corresponding product, thus implicitly demonstrating the familiarity of the names and of the metonymic transfer, although not fully lexicalized. For approximation, data suggested that the selected adjectives were judged appropriate when referred to the nouns used in the approximations, yet less typical than when used literally, providing first quantitative evidence for the definition of the category of approximation.

The set of stimuli also appears to be well controlled for the contextual expectancy: cloze probability was very low across sets, never above 12.00% for any condition of any sets. Interestingly, participants never created metaphor nor approximation in completing the sentences. Strictly speaking, also for metonymy the cloze probability was equal to zero. However, we observed 7 cases in which the final word reported by the participants was a proper name, albeit different from the one used in the corresponding stimulus (e.g., original stimulus: ‘That writer translates Fruttero’; cloze probability results: ‘That writer translates... Hesse / Sartre’). These results highlighted that there were some verbs spontaneously used in their metonymic sense, while the probability of creating a metaphor or an approximation was not verified, suggesting that metonymy is somehow more prone to routinization.

As a final note, one may argue that target words had different syntactic functions across sets, being used predicatively in metaphor and approximation, and referentially in the case of metonymy. However, we believe that it was important to focus on standard uses of the three pragmatic phenomena, rather than maintaining the same target word at the price of less clear and prototypical pragmatic types.

3 Timed sensicality judgment study

Timed sensicality judgments task has been used as a valuable paradigm to explore interpretation assignment, at different levels of the linguistic structure. This paradigm has been widely employed in investigations targeting conceptual operations, including polysemy (Klein & Murphy 2001) and compounds (Gagné 2001). At the sentence level, sensicality judgments have been used to explore pragmatic interpretation of conjunctions (Bott et al. 2009). The advantage of the sensicality judgment task is that it requires not only to access but also to elaborate the meaning of the expression. Information can be gathered both on the availability of the correct interpretation under time pressure (measured in terms of accuracy, i.e., proportion of correct responses - judging a sensible expression to be sensical or a nonsense expression to be non-sensical) and on the costs of interpretation (measured in terms of latencies). Interestingly, sensicality judgments often recur as task in a number of experimental paradigms targeting figurative language processing, from Speed-Accuracy Tradeoff (McElree & Nordlie 1999) to neurophysiological and neuroimaging studies (Arzouan et al. 2007; Lai et al. 2009; Rapp et al. 2011; Subramaniam et al. 2012).

Here we used timed sensicality judgments to explore interpretation assignment of different types of pragmatic uses compared to literal counterparts in order to test (i) the Relevance Theory claim that metaphor and approximation are based on the same kind of conceptual adjustment yet in different degree and directions, and (ii) the hypothesis – to a certain extent shared by Relevance Theory, Cognitive Linguistics and other frameworks – that metonymy is supported by conceptual processes different from those involved in metaphor processing. If the first hypothesis is correct, and specifically approximation involves only a marginal broadening while metaphor involves a wider broadening (Wilson 2003) or broadening coupled with narrowing (Carston & Wearing 2011), then the interpretation of metaphor and approximation should be similar but with a gradient of costs: both metaphor and approximation should depart from literal interpretation, but they should do so differently, with metaphorical interpretation less available and more difficult (lower accuracy) and associated with higher costs (higher latencies) than approximation. If the second hypothesis is correct, and specifically metaphor and metonymy differ in terms of underlying conceptual operations, the latter not requiring broadening and narrowing but rather conceptual shift, different patterns in availability and costs are expected. In particular, based on previous experimental evidence, it is possible that metonymic

interpretation doesn't come with extra cost with respect to literal comprehension, due to the routinization of the transfer type.

3.1 Methods

Participants

Twenty-five native speakers of Italian (12 M /13 F; mean age = 25.32 ± 3.02 years; mean schooling years = 18.3 ± 3.03) participated in the study. Participants were unaware of the aim of the study, and not experts in linguistics or psycholinguistics. None of them had participated in the rating study. They gave written consent to participate after receiving an explanation of the procedures, according to the Declaration of Helsinki, and received a monetary reimbursement for their participation.

Materials

The final pool of sentences described in the previous section was used as stimuli, i.e., 42 triplets for each set (Metaphor, Metonymy, Approximation). Additional 42 anomalous sentences were included for each set, with the purpose of having a similar ratio of sense and non-sense items. In order to minimize potential effects related to the repetition of the target words X in the triplets, the additional items recombined other words in the set, partly by repeating the subject nouns Y (e.g., 'Those insects are tables', where 'insects' is the subject noun in the literal version of one triple in the metaphor set; see Table 1) and partly by repeating the last word of the additional item (e.g., 'Those trousers are tables', where 'tables' is the last word of the additional item obtained as above). Furthermore, to reduce the proportion of pragmatically used words and avoid metalinguistic awareness on figurative language, the experimental items were intermixed with 594 fillers (66% sense, 33% non-sense), consisting of four word sentences, like the experimental stimuli. In total, there was a sense:non-sense ratio of 1.44:1, and pragmatic sentences represented 12.0% of the stimuli (4.0% metaphors, 4.0% metonymy, and 4.0% approximation).

Procedure

Each participant was tested individually. Stimulus presentation and response collection were all carried out on a personal computer, using Presentation© software.

Each trial began with a fixation cross presented in the middle of the screen for 500 ms. Next, the sentence was presented word by word at fixed rate (300 ms). After the final word, YES/NO appeared on the screen to indicate that participants could give their response. Participants were instructed to respond as quickly and accurately as possible, and to make a sensicality judgment by pressing the green button when the string was meaningful and the red button when the sentence was meaningless. The assignment of red and green to the left and right keys was counterbalanced across participants. After response or time-out (4000 ms), there was a blank inter-trial interval of 1000 ms. Response times were measured from the offset of the target word X.

Each subject was presented with all sentences. To avoid fatigue, three experimental blocks were created. An equal number of pragmatic, literal, and anomalous sentences from each set were included in each block, along with an equal number of fillers. We assigned the members of each triplet and the additional anomalous counterpart to distinct blocks, in order to avoid long-distance priming effects. Within each block, sentences were presented in a random order, while the order of the block was pseudo-randomized across participants. Mandatory stops between experimental blocks were fixed. A training session including ten items preceded the experiment. Furthermore, two practice trials (not included in the analysis) were administered at the beginning of each block. Overall the experimental session lasted 1 hour.

3.2 Results

	Metaphor set		Metonymy set		Approximation set	
	Accuracy	Reaction time	Accuracy	Reaction time	Accuracy	Reaction time
Pragmatic	0.52 (0.24)	744.66 (339.02)	0.83 (0.10)	688.70 (338.17)	0.87 (0.09)	655.96 (308.82)
Literal	0.95 (0.04)	638.10 (305.05)	0.89 (0.07)	658.32 (326.55)	0.95 (0.04)	591.13 (293.18)
Anomalous	0.97 (0.05)	653.98 (339.44)	0.95 (0.06)	669.01 (332.04)	0.85 (0.03)	673.37 (333.20)

Table 3. Accuracy rates and mean reaction times (ms) for correct responses as a function of pragmatic modulation (pragmatic, literal, anomalous conditions) and set type (Metaphor set, Metonymy set, Approximation set). Standard deviations in parentheses.

Responses faster than 250 ms and slower than 1750 ms were excluded from the analysis (10.4 % of the data). We also excluded data by two participants with overall accuracy rate lower than 80%, and by one participant with 40% of responses faster than 250 ms. Table 3 presents accuracy rates and mean reaction times for correct responses for each experimental conditions.

Accuracy

A Univariate General Linear Model (PASW Statistics 18.0.0) with pragmatic modulation (three levels: pragmatic, literal, anomalous) and set type (metaphor set, metonymy set, approximation set) as fixed factors was carried out on accuracy rates, treating either subjects (F1) or items (F2) as a random factor. Results showed that both the pragmatic modulation factor (F1 (2,168) = 60.74, $p < 0.001$; F2 (2,328) = 68.98, $p < 0.001$) and the set type factor (F1 (2,168) = 10.74, $p < 0.001$; F2 (2,328) = 12.41, $p < 0.001$) were significant. Also their interaction was significant (F1 (4,168) = 32.33 $p < 0.001$; F2 (4,328) = 37.47, $p < 0.001$), indicating that the effect of one factor depends on the level of the other factor.

We therefore explored the effect of pragmatic modulation set by set, focusing on the comparison between the pragmatic and the literal sentences. In the Metaphor set, this factor yielded significant effects (F1 (2,42) = 68.43, $p < 0.001$; F2 (2,82) = 100.77, $p < 0.001$), with metaphorical sentences less accurate than literal sentences (Tukey's HSD, $p < 0.001$ both by subjects and by items). Also in the Metonymy set pragmatic modulation was significant (F1 (2,42) = 12.73, $p < 0.001$; F2 (2,82) = 8.10, $p = 0.001$). Differing from the Metaphor set, however, accuracy doesn't seem to vary for metonymic and literal sentences: post-hoc comparisons revealed no difference between the two conditions (Tukey's HSD, $ps \geq 0.05$). In the Approximation set, again we observed a main effect of pragmatic modulation (F1 (2,42) = 3.94, $p = 0.02$; F2 (2,82) = 8.49, $p = 0.001$). Accuracy for approximation was significantly lower than for literal sentences in the by item analysis (Tukey's HSD, $p = 0.006$), although the difference was not significant in the by subject analysis (Tukey's HSD, $p = 0.10$).

Overall, the data suggest a higher availability of literal uses as compared to metaphor and approximation, but not for metonymy.

Latencies

Following the standard in analyzing response times, only trials in which participants responded correctly were included in the analysis. The effect of pragmatic modulation (three levels: pragmatic vs. literal vs. anomalous) and set type (three levels: Metaphor set, Metonymy set, Approximation set) on response times were examined with Univariate General Linear Model treating either subjects (F1) or items (F2) as a random factor. We observed a significant effect of pragmatic modulation (F1 (2, 167) = 12.23, $p < 0.001$; F2 (2, 327) = 24.97, $p < 0.001$), as well as a significant effect of set type in the by item analysis and marginally significant in the by subject analysis (F1 (2, 167) = 2.95, $p = 0.05$; F2 (2, 327) = 9.46, $p < 0.001$). A significant interaction between pragmatic modulation and set type was found (F1 (4, 167) = 3.77, $p = 0.006$; F2 (4, 327) = 8.70, $p < 0.001$), as shown in Figure 2.

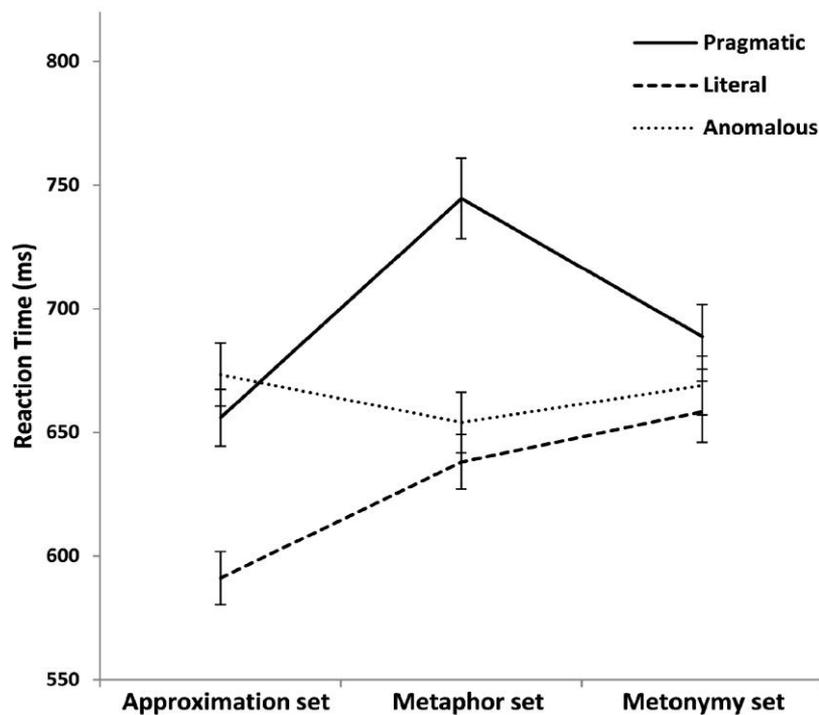


Figure 2. Mean reaction times (ms) for the Approximation set, the Metaphor set and the Metonymy set as a function of the pragmatic modulation factor. Pragmatic level is represented by the solid line, literal level by the dashed line, and anomalous level by the dotted line. Error bars indicate standard error.

In order to explore the interaction of pragmatic modulation and set type, simple effect analyses were conducted. As concerns the pragmatic modulation factor, in the Metaphor set we observed that metaphorical sentences were interpreted slower than literal counterparts ($F_1(2, 41) = 5.85, p = 0.006$; $F_2(2, 81) = 25.51, p < 0.001$; Tukey's HSD, $p < 0.01$ both by subjects and by items)³. On the contrary, in the Metonymy set there were no differences across conditions ($F_1(2, 42) = 0.74, p = 0.48$; $F_2(2, 82) = 2.06, p = 0.13$), indicating that metonymic interpretation was reached as rapidly as literal interpretation. Similarly to the Metaphor set, the Approximation set showed a significant effect of the pragmatic modulation factor ($F_1(2, 42) = 10.66, p < 0.001$; $F_2(2, 82) = 8.37, p < 0.001$), with approximations interpreted slower than literal sentences (Tukey's HSD, $p < 0.01$ both by subjects and by items).

We also assessed whether the type of pragmatic use has an effect on response times. Since pragmatic sentences differ in some respects – as needed to preserve clear pragmatic types, we avoided direct comparisons of metaphors, metonymies and approximations across sets. Rather, we measured the latency difference between the pragmatic condition and the literal condition for the corrected pairs of each set: (metaphor – literal), (metonymy – literal) and (approximation – literal), as represented in Figure 3. The highest latency difference was obtained for metaphor (M metaphor – literal = 105.34 ms), followed by approximation (M approximation – literal = 78.50 ms), while a minimal latency difference was observed for metonymy (M metonymy – literal = 2.76 ms). The comparison reveals an effect of the type of pragmatic use ($F(2, 41) = 14.14, p < 0.001$), with metaphor and approximation significantly different from metonymy, but not different from each other (Tukey HSD: metaphor/approximation vs. metonymy, $p \leq 0.001$; metaphor vs. approximation, $p = 0.56$).

³ Degrees of freedom vary for the Metaphor set as compared to the other sets due to one participant who never answered correctly for metaphors (not excluded from the analysis as his overall level of accuracy was higher than the 80% threshold) and due to one metaphorical item that was never judged accurately (not excluded from the analysis based on the results of the rating study).

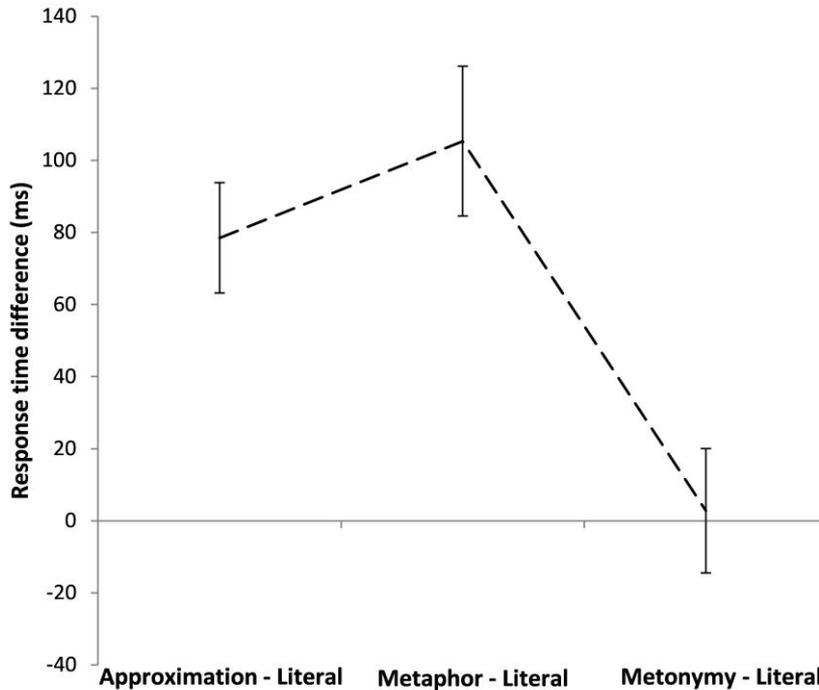


Figure 3. Reaction times differences (pragmatic minus literal) for the Approximation set, the Metaphor set and the Metonymy set. Error bars correspond to the standard error of the difference for each set.

4 General discussion

The results of the rating study indicated that the three pragmatic uses were easily interpreted as meaningful in a manner similar to the corresponding literal constructions. However, timed sensicality judgments revealed that there are differences across metaphor, metonymy and approximation, reflected both in accuracy rates and latencies of interpretation.

First, pragmatic modulation affects the accuracy of the response. Assigning a pragmatic interpretation under the pressure of time seems to be more difficult when we depart from literal meaning. There are, however, notable differences. For metaphor, the percentage of correct responses was around 50%. Although judged as meaningful in the offline rating, in timed sensicality judgments metaphors proved more difficult to interpret. This piece of evidence is consistent with previous literature

employing sensicality judgments, from Speed Accuracy Tradeoff (McElree & Nordlie 1999) to neurophysiological recording (Arzouan et al. 2007) and neuroimaging (Subramaniam et al. 2012). Accuracy around chance seems thus a common performance associated to metaphor interpretation in speeded condition. Crucially, participants performed much better when they were presented with approximations and metonymies, reaching 87% and 85% respectively. This suggests higher availability for approximate and metonymic uses with respect to metaphorical use. Interestingly, we observed that metaphors and approximations tended to be interpreted less accurately than their literal counterparts, while there are no accuracy differences between metonymies and literal expressions. This points in the direction of similarities in the interpretation style of metaphor and approximation, although associated with different degrees of availability, while the availability of metonymy seems to equate that of literal interpretation⁴.

Second, also the speed at which pragmatic interpretation is reached is influenced by pragmatic modulation. Analyzing error-free trials, that is limiting the analysis to those cases where pragmatic sentences were judged to be sensical, we observed that interpreting pragmatic uses is not always slower than literal interpretation. For metaphor and approximation this is the case, while for metonymy it is not. These findings accord with a vast behavioral literature on metaphor showing that, in minimal context and for not conventionalized expressions, metaphor processing requires extra costs compared to literal processing (Noveck et al. 2001). Here, we carefully controlled the sentential environment, by providing each pragmatic use with a minimal context, and the familiarity of the expressions, by avoiding lexicalized cases. Thus, the higher reaction times for metaphor seem to truly reflect extra costs required by the interpretation of metaphorically used words as compared to literal uses. Results are also consistent with eye-tracking studies reporting no differences between metonymic and literal expressions. It cannot go unnoticed, however, that, at the neural level, metonymic expressions elicit robust differences from literal comprehension

⁴ Note that in the Metonymy set literal sentences reached only 89% accuracy. This is probably related to the costs of processing proper names, which require the retrieval of stored knowledge, also when they are well-known by the participants (as assessed in the rating study). However, this should not affect the comparison between metonymies and literal controls, as proper names were included in both conditions and their presence should not block interpretative differences to emerge, if any. On the demands placed on retrieval of stored knowledge for proper names as compared to common names, see Gorno-Tempini et al. (1998).

(Schumacher 2011; Rapp et al. 2011). It is up to future studies to elucidate whether this discrepancy is motivated by differences in the materials, either in conventionality or supportive context, or in the methodological techniques and the type of information they offer. Our view is that, when metonymy – like in our case – is based on common shifts such as producer for product, no matter the relative conventionality of the specific lexical items, in a minimal yet sufficient context, interpretation costs in speeded conditions closely mirror literal comprehension, and radically differ from those required by metaphor processing. Besides, this view is consistent with findings reported for other types of routinized meaning shift. For example, sensicality judgments on logical metonymy showed no differences in accuracy nor in latencies between the coerced and control conditions, yet again evoking neural differences (Brennan & Pylkänen 2008).

Reexamining previous studies targeting directly the metaphor/metonymy distinction, our findings appear well compatible with the acquisition data, which showed that metonymy not only is acquired at a faster rate than metaphor, but it also processed more accurately throughout childhood to adulthood. The authors hypothesized a more basic type of conceptual operation for metonymy as opposed to metaphor, reflected in the reduced difficulty and costs observed here (Rundblad & Annaz 2010), which is compatible with both the relevance theoretic and cognitive linguistic view. Furthermore, in some cases metonymic meanings might be part of the lexicon, as suggested by the correlation of metonymic comprehension with the expansion of receptive vocabulary (Annaz et al. 2009).

It might be of some interest here to report some qualitative insights from the post-experiments session: despite the very low percentage of metaphors in the sentence pool, some participants noticed their presence, while none seemed to notice metonymy, as if metonymic uses were more integrated in the lexical knowledge and less prominent in the speakers' metalinguistic awareness. By contrast, our results seem to be conflicting with the higher reading times for resolving metonymic referring expressions as compared to metaphorical referents presented in Gibbs (1990). However, those data are controversial, as potentially affected by the plausibility of the items (Frisson & Pickering 1999), and obscuring some comparisons of interest (Noveck et al. 2001).

The different behavior observed for metaphor and metonymy gains support from the results on approximation due to the theory-grounded test-bed they offer. We

showed that adjectives used approximately are interpreted slower than the same adjectives used literally. This piece of evidence seems to place approximation closer to metaphor, as Relevance Theory would predict, than to metonymy. Furthermore, this result seems to strengthen the distinction between the processing styles for metaphor and for metonymy, by introducing a third case that patterns alike the former but differently from the latter. Consonant with this are also the latency differences between the pragmatic and the literal conditions across sets. When we disentangle the costs of interpreting each type of pragmatic use, we see that approximation and metaphor are associated with extra costs, while metonymy doesn't prompt extra effort. Accuracy data, with approximation and metaphor departing from literality, and metonymy equating it, are in harmony.

Collectively, this pattern of results carries importance for discussing theoretical accounts of the nature of pragmatic phenomena. The relevance-theoretic claim that metaphor and approximation both require conceptual adjusting of the linguistically encoded concept but in different degrees (Wilson 2003) seems to be supported by our data, and specifically by the gradient observed in availability and latency. Also the direction of the extension may be the cause of the different gradient observed in the sensicality judgments.

In this light, following Carston & Wearing (2011) on hyperbole and extending to approximation, the different difficulty and costs might stem from marginal broadening in the case of approximation as opposed to broadening coupled with narrowing in the case of metaphor. In uses such as 'Those tires are smooth', the hearer must derive an ad hoc concept SMOOTH* which is more general than the encoded concept, in order to include also surfaces that do not possess all the sense-related properties of smoothness, but the widening is minimal. Conversely, in 'Those dancers are butterflies', BUTTERFLIES* is an ad hoc modification that radically broadens the denotation of the lexical concept, in order to include perhaps graceful entities, and also narrows down the denotation to exclude, for instance, other elegant flying insects. For the type of task used here, we do not have direct evidence to discriminate whether the difference between approximation and metaphor lays in marginal versus radical broadening or in marginal broadening versus a combination of radical broadening and narrowing. Intuitively, our data fit well with the degree claim posited by Relevance Theory, while the direction claim is less straightforwardly answerable. More sophisticated designs will be needed that manipulate the (degree of) direction of the adjustment, possibly exploring the temporal dynamics of the process or the conceptual

properties that undergo manipulation. Granted this caveat, the general idea of a modulation in the underlying conceptual adjustment process seems to be well supported by our findings.

Converging evidence comes from Deamer et al.'s (2010) reading time study, where hyperbolic uses (e.g., 'The back yard definitely needed pruning. It was a forest') were compared to metaphorical uses (e.g., 'Sam always got lost. The university was enormous. It was a forest'), showing that even a more substantial types of broadening such as hyperbole is distinct from metaphor. This study actually failed in finding a difference between hyperboles and literal expressions. Far from finding this result in conflict with our data, we believe that contextual modulation is the cause of the discrepancy: Deamer et al. used supportive contexts (as in the examples above, and including also a condition with longer passages) that might have facilitated hyperbole resolution and reduced the broadening, while we used a minimal sentential environment that allowed for the marginal extra costs required by approximation to emerge.

Also the hypothesis that metonymy is contingent on a different interpretative style, not straightforwardly reducible to narrowing or broadening but involving some kind of shift, fits with our data. The different pattern of results observed for metonymy as opposed to metaphor might reflect different conceptual operations. Going further, our data also point to reduced efforts for metonymy, and to the routinization of some types of metonymic shifts, such as producer for product. A possible distinction might be sketched between the combination of broadening and narrowing on the metaphor side, less pre-configured in direction and degree of the conceptual adjustment of the lexical concept, and conceptual shift on the metonymy side, based on more routinized patterns. Highly creative metonymic uses are possible as well (consider, for instance, "The best pencils of the world gather together for the annual drawing convention"), and this might call upon higher interpretation costs. However, it seems psychologically implausible to posit different elaboration procedures for the same class of phenomena, as the difference between routinized producer for product cases and less typical tool for worker cases could probably be made not by different types of conceptual adjustment processes, but rather by the role of context.

The results described for metonymy can also be reconciled with the Cognitive Linguistics account, to the extent that metaphor and metonymy are ascribed to distinct types of mappings: across domains for metaphor, within the same domain for

metonymy. A greater cognitive distance between concepts can be assumed for metaphor (Rundblad & Annaz 2010) and might be reflected in higher difficulty and costs. As it is still difficult to translate the different types of mappings in terms of processing costs, we leave this for further research to develop. Other explanations are also possible. Differences might not be confined to the conceptual level, and involve the derivation of implicatures. Metaphor is well known for generating a wide array of weak implicatures, while the implicatures associated with metonymy are less explored (but see Papafragou 1996). It has also been shown that the inferential chain related to the speech act type is a primary element in modulating the processing costs and marking the difference between figurative types (Bosco et al. 2009).

All these possible explanations, however, need to be further investigated in future studies that could enable us to tap into earlier processing stages, and to unravel conceptual operations and manipulations. Sensicality judgments are a good measure of the availability and difficulty of correct interpretation, but are limited to stages where the sense has already been construed, and do not account for online processing (Frisson 2009). Thus, our results shed light on the costs of interpretation assignment, and provide insights into interpretative style, but the temptation of discussing issues such as the temporal dynamics and the deep nature of the conceptual operations must be resisted for the moment.

5 Conclusions

Behind a label such as figurative language, many different mechanisms are grouped. Although we assume that all require pragmatic inferencing to be interpreted, interpretation might come with different procedures, linked to different operations at the conceptual level. Through timed sensicality judgments recorded for different pragmatic uses in minimal context condition, we found that there are significant differences in the interpretation availability and costs of metaphor, metonymy and approximation. The findings support a theoretical distinction between metaphor and approximation, which seem to vary in degree and possibly in the direction of the underlying adjustment process, as predicted by Relevance Theory, and an even more marked separation with metonymy, whose meaning shift might be subject to routinization. With these data, we hope to have strengthened the empirical basis available on figurative language, by providing the first evidence in favor of the psychological reality of the phenomenon of approximation, and with a first attempt to

answer the challenge raised by metonymy. We believe that deepening the understanding of the phenomena included under the realm of pragmatics, by pinpointing potential differences for the parser and elaborating on whether natural classes of cases can be identified on this basis, is one promising line of research for the experimental pragmatics enterprise.

References

- Annaz, Dagmara, Jo Van Herwegen, Michael Thomas, Roza Fishman, Annette Karmiloff-Smith & Gabriella Rundblad (2009), Comprehension of metaphor and metonymy in children with Williams syndrome. *International Journal of Language & Communication Disorders* 44(6): 962–978.
- Arzouan, Yossi, Abraham Goldstein & Miriam Faust (2007), Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension. *Brain Research* 60: 69–81.
- Bambini, Valentina, Claudio Gentili, Emiliano Ricciardi, Pier Marco Bertinetto & Pietro Pietrini (2011), Decomposing metaphor processing at the cognitive and neural level through functional magnetic resonance imaging. *Brain Research Bulletin* 86: 203–216.
- Bertinetto, Pier Marco, Cristina Burani, Alessandro Laudanna, Lucia Marconi, Daniela Ratti, Claudia Rolando & Anna Maria Thornton (2005), *Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS)*, <http://linguistica.sns.it/CoLFIS/Home.htm>
- Bosco, Francesca M., Marianna Vallana & Monica Bucciarelli (2009), Comprehension of communicative intentions: The case of figurative language. *Journal of Cognitive Science* 10: 245-277.
- Bott, Lewis, Steven Frisson & Gregory L. Murphy (2009), Interpreting conjunctions. *Quarterly Journal of Experimental Psychology* 62: 681–706.
- Brennan, Jonathan & Liina Pylkkänen (2008), Processing events: Behavioral and neuromagnetic correlates of Aspectual Coercion. *Brain & Language* 106: 132–143.
- Cacciari, Cristina & Sam Glucksberg (1994), Understanding figurative language. In Morton A. Gernsbacher (ed.), *Handbook of Psycholinguistics*, New York: Academic Press: 447–477.
- Cardillo, Eileen R., Gwenda L. Schmidt, Alexander Kranjec & Anjan Chatterjee (2010), Stimulus design is an obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavior Research Methods* 42: 651–664.
- Carifio, James & Rocco Perla (2008), Resolving the 50-year debate around using and misusing Likert scales. *Medical Education* 42(12): 1150–1152.

- Carston, Robyn (2009), Explicit/implicit distinction. In Louise Cummings (ed.). *The Pragmatics Encyclopedia*. London: Routledge.
- Carston, Robyn (2010), Lexical pragmatics, ad hoc concepts and metaphor: A Relevance Theory perspective. *Italian Journal of Linguistics/Rivista di Linguistica* 22(1): 153–180.
- Carston, Robyn & Catherine Wearing (2011), Metaphor, hyperbole and simile: A pragmatic approach. *Language and Cognition* 3(2): 283–312.
- Conover, W. Jay, & Ronald Iman (1981), Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician* 35: 124–129.
- De Grauwe, Sophie, Abigail Swain, Philip J. Holcomb, Tali Ditman & Gina R. Kuperberg (2010), Electrophysiological insights into the processing of nominal metaphors. *Neuropsychologia* 48: 1965–1984.
- Deamer, Felicity, Nausicaa Pouscoulous & Richard Breheny (2010), A contrastive look at metaphor and hyperbole. *UCL Working Papers in Linguistics* 22.
- Egg, Markus (2004), Metonymie als Phänomen der Semantik-Pragmatik-Schnittstelle. *Metaphorik* 6: 36–53.
- Frisson, Steven (2009), Semantic underspecification in language processing. *Language and Linguistics Compass* 3(1): 111–127.
- Frisson, Steven & Martin J. Pickering (1999), The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25(6): 1366–1383.
- Frisson, Steven & Martin J. Pickering (2007), The processing of familiar and novel senses of a word: Why reading Dickens is easy but reading Needham can be hard. *Language and Cognitive Processes* 22: 595–613.
- Gagné, Christina L. (2001), Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27(1): 236–254.
- Gibbs, Raymond W. (1990), Comprehending figurative referential descriptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16: 56–66.
- Gibbs, Raymond W. (1994), *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge: Cambridge University Press.

- Giora, Rachel (2003), *On our Mind: Salience, Context, and Figurative language*. New York: Oxford University Press.
- Gorno-Tempini, Maria L., Cathy J. Price, Oliver E. Josephs, Rik Vandenberghe, Stefano F. Cappa, Narinder Kapur & Richard S. Frackowiak (1998), The neural systems sustaining face and proper-name processing. *Brain* (121): 2103–2118.
- Greenacre, Michael (1993), *Correspondence Analysis in Practice*. London, Academic Press.
- Grice, H. Paul (1975), Logic and conversation. In Peter Cole and Jerry L. Morgan (eds.). *Syntax and Semantics: Speech Acts*. New York: Academic Press.
- Jamieson, Susan (2004), Likert scales: How to (ab)use them. *Medical Education* 38(12): 1217–1218.
- Kintsch, Walter & Anita Bowles (2002), Metaphor comprehension: What makes a metaphor difficult to understand? *Metaphor and Symbol* 17: 249–262.
- Klein, Devorah E. & Gregory L. Murphy (2001), The representation of polysemous words. *Journal of Memory and Language* 45: 259–282.
- Knapp, Thomas R. (1990), Treating ordinal scales as interval scales: an attempt to resolve the controversy. *Nursing Research* 39(2): 121–123.
- Lai, Vicky T., Tim Curran & Lise Menn (2009), Comprehending conventional and novel metaphors: An ERP study. *Brain Research* 11: 145–155.
- Lakoff, George & Michael Johnson (1980), *Metaphors We Live By*. Chicago: University of Chicago Press.
- Lascarides, Alex & Ann Copestake (1998), Pragmatics and word meaning. *Journal of Linguistics* 34(2): 387–414.
- Mcelree, Brian & Johanna Nordlie (1999), Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review* 6(3): 486–494.
- Norman, Geoff (2010), Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education: Theory and Practice* 15(5): 625–632.
- Noveck, Ira A., Maryse Bianco & Alain Castry (2001), The costs and benefits of metaphor. *Metaphor and Symbol* 16(1&2): 109–121.
- Noveck, Ira A. & Anne Reboul (2008), Experimental Pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences* 12: 425–431.

- Nunberg, Geoffrey (1995), Transfers of meaning. *Journal of Semantics* 12: 109-32.
- Panther, Klaus Uwe & Linda Thornburg (2007), Metonymy. In Dirk Geeraerts and Hubert Cuyckens (eds.), *The Oxford Handbook of Cognitive Linguistics*. Oxford, OUP: 236–263.
- Papafragou, Anna (1996), On metonymy. *Lingua* 99: 169–195.
- Pustejovsky, James (1995), *The Generative Lexicon*. Cambridge, MA: MIT Press.
- Rapp, Alexander, Michael Erb, Wolfgang Grodd, Mathias Bartels & Katja Markert (2011), Neural correlates of metonymy resolution. *Brain and Language* 119(3): 196–205.
- Recanati, François (2010), *Truth-conditional Pragmatics*. Oxford: Oxford University Press.
- Ruiz de Mendoza Ibáñez, Francisco J. (2007), High-level cognitive models: In search of a unified framework for inferential and grammatical behavior. In Krzysztof Kosecki (ed.) *Perspectives on Metonymy*. Frankfurt/Main. Peter Lang: 11-30.
- Rundblad, Gabriella & Dagmara Annaz (2010), Development of metaphor and metonymy comprehension: Receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology* 28: 547–563.
- Schumacher, Petra B. (2011), The hepatitis called ...: Electrophysiological evidence for enriched composition. In Jörg Meibauer and Markus Steinbach (eds.) *Experimental Pragmatics/Semantics*. Amsterdam/Philadelphia: John Benjamins: 199–219.
- Schumacher, Petra B. (2012), Context in neurolinguistics: Time-course data from electrophysiology. In Rita Finkbeiner, Jörg Meibauer and Petra B. Schumacher (eds.), *What is a Context? Linguistic Approaches and Challenges*. Amsterdam/Philadelphia: John Benjamins: 33–53.
- Sperber, Dan & Deirdre Wilson (2005), Pragmatics. In Frank Jackson and Michael Smith (eds.), *Oxford Handbook of Contemporary Philosophy*. Oxford: Oxford University Press: 468–501.
- Subramaniam, Karun, Miriam Faust, Mark Beeman & Nira Mashal (2012), The Repetition Paradigm: Enhancement of novel metaphors and suppression of conventional metaphors in the left inferior parietal lobe. 10.1016/j.neuropsychologia.2012.07.020.

- Wilson, Deirdre (2003), *Relevance Theory and lexical pragmatics*. *Italian Journal of Linguistics/Rivista di Linguistica* 15: 273–291.
- Wilson, Deirdre & Robyn Carston (2006), *Metaphor, relevance and ‘the emergent property’ issue*. *Mind & Language* 21: 404–433.
- Wilson, Deirdre & Robyn Carston (2007), *A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts*. In Noel Burton-Roberts (ed.), *Pragmatics*, Palgrave Macmillan: Basingstoke and New York: 230–259.