Lexical ambiguity in elementary inferences: 
an experimental study

Francesca Ervas, Elisabetta Gola, Antonio Ledda, Giuseppe Sergioli
(University of Cagliari)

Abstract

In this paper we discuss how common meaning ambiguities (homonymy, polysemy and metaphors) influence the understanding of an elementary argument. In order to understand how, and to what extent, participants’ intuitions on the strength of a syllogism are influenced by meaning ambiguities, we present the results of a pilot study. The study specifically focuses on a fallacy of lexical ambiguity, where the meanings of the middle term diverge in the two premises. We hypothesize that the evaluation of the strength of an argument of this sort is related to the nature of the ambiguity of its middle term and to the pragmatic process required to disambiguate the ambiguous meanings. We expect the persuasiveness of the syllogism to be directly proportional to the degree of semantic superposition of the meanings of the middle term.

1. Introduction

In the informal logic tradition, fallacies are common errors of reasoning (Sergioli 2015). As it has been pointed out, «studies of fallacies in argumentation and informal logic have mainly taken a normative approach, by seeing fallacies as arguments that violate standards of how an argument should properly be used in rational thinking or arguing. However, fallacies also have a psychological dimension» (Walton 2010, p. 159, our italics). In this perspective, fallacies are arguments that seem valid but are not (Hamblin 1970). In Max Black’s Critical Thinking, the definition of fallacy is tuned up: «A fallacy is an argument that seems to be sound without being so in fact. An argument is “sound” for the purpose of this definition if the conclusion is reached by a reliable method and the premises are known to be true» (Black 1946, pp. 129-130, our italics). However little is said in a manner of explanation regarding the how and why of “seeming valid” or “seeming sound” (Hansen 2002). Introducing lexical fallacies, Cohen and Nagel suggested that «these all seem to conform to valid forms of inference, but on careful examination are seen not to do so the appearance being due to an ambiguity, that is, to the use of the same word or verbal sign for two different terms» (1934, p. 376) .
In classical argumentation theory, metaphors usually lead to fallacies of reasoning (Sergioli & Ternullo 2014; Tindale 2006), by exacerbating problems of ambiguity (Barnden 2012; Camp 2006; Fischer 2015). A sentence where a metaphor occurs is literally false. However, the context of its usage might create a perception that it is true, or at least plausible. From a literal point of view a metaphor is false, but from a non-literal point of view it may appear plausible. Fischer (2015) argues that a range of varied disciplines have demonstrated the productive usage of metaphors in deductive reasoning: physics (Hesse 1996), biology (Keller 1995), psychology (Gentner & Grudin 1985) and problem solving (Keefer et al. 2014; Thibodeau & Boroditsky 2011). At the same time, metaphors are governed by heuristic rules that never guarantee the preservation of truth, thus giving rise to systematic fallacies (Fischer 2011, 2014). This might be the reason why metaphors are highly persuasive by nature. Thus, the persuasiveness of an argument would eventually be influenced by the presence of metaphors.

Following this intuition, we aim to understand the effect of metaphors in arguments with the structure of a quaternio terminorum, sometimes referred to as four term fallacy, where the nature of the middle term plays a fundamental role and might influence its persuasive strength. Quaternio terminorum relies on the intrinsic ambiguity of the middle term, which is used in the premises with two different meanings. The middle term can be ambiguous either because the term has two different literal meanings (the case of homonymy and polysemy), or because it has both a literal and a non-literal meaning (the case of lexicalized metaphor and live metaphor). We claim that the persuasiveness of an argument varies as the ambiguity of the middle term moves through such a spectrum.

In order to test our claim on a real representative sample, we set up an experimental design. Usually, argumentation theory does not employ an experimental approach and relies on a priori notions, according to which figurative language renders premises literally false and the overall argument unsound. On the contrary, an experimental approach may tell us something more about argumentation in common language and the divergences between speakers’ and scholars’ intuitions on the strength of an argument; in particular, on the reason why a fallacious argument may appear strong, even if it is not. In particular, the results of this study indicate a consistency between the psychological and the normative aspects (Walton 2010; Tversky & Kahneman 1974) of quaternio terminorum. The semantic intuitions leading the disambiguation of the middle terms allow participants to correctly evaluate the arguments. This is so with participants bearing no specific training in logic nor in argumentation theory. Interestingly, this consistency seems to weaken in case of metaphors as middle terms. This is perhaps due to the fact that the presence of metaphors renders natural language ambiguous.

The structure of the paper is the following. In the first section we provide the basic notions required to define quaternio terminorum and clarify the different pragmatic phenomena involved in the gamut of lexical ambiguities. In the second section, we present the experimental design of a pilot study that aims to understand whether, and to what extent, lexical ambiguity influences the persuasiveness of a quaternio terminorum. In the third section, we present the methodology of the experimental study, and describe the participants involved, either in the norming or the pilot study, the materials presented to the participants, and the procedure. In the

---

1 In this paper we will be exclusively concerned with literally false metaphors. However, it should be observed that there are cases of twice-true (Cohen 1976, p. 254) and twice-apt (Hills 1997, p. 130) metaphors, such as “No man is an island” and “George Bush is a primate”. See Camp (2008) for an extensive discussion.
fourth section we present the results of the experimental study. Finally, in the fifth section, we discuss the results in the light of the aims of the overall paper.

1.1. Basic notions

In order to keep the paper self-contained, we introduce in this section all the required basic notions from argumentation theory and lexical pragmatics.

1.1.1 Argumentation theory

A sentence is a declarative statement to which a truth-value (true or false), in classical logic, can be assigned (Govier 1999; Copi & Cohen 2014). By an argument or a syllogism we mean a set of three sentences: two premises (major and minor, respectively) and a conclusion, which involves three terms: the subject, the predicate of the conclusion, and a third term (middle term) connecting the subject of the first premise to the predicate of the second.

Argumentation theory discusses the conditions under which a conclusion follows from true premises. To this effect, the concepts of validity and soundness are required (Walton 2005, 2010):

- an argument is **valid** if its conclusion is true, whenever its premises are true;
- an argument is **sound** if it is valid and all its premises are true.

Validity and soundness are crucial in evaluating arguments. Nonetheless, these notions seem too tight for ordinary communication. Due to this reason, two weakened versions of these concepts may be expedient: **strength** and **goodness** (Bonissone 1987; Borwein & Bailey 2008; Epstein & Kernberger 2006, Ch. 3).

- an argument is **strong** if it is very likely that its conclusion is true, whenever its premises are true. In other words, premises provide reasons that support the probable truth of the conclusion. We call weak an argument which is not strong.
- an argument is **good** if it is strong and all its premises are plausible.

In this article we will deal with the strength and the goodness of those arguments whose middle term potentially call for (at least) two distinct meanings, as mentioned in the introduction. Quaternio terminorum (Copi & Cohen, 2014, pp. 230-231; Smiley 1973, pp. 136-154) is a well known case of a fallacious argument based on the ambiguity of its middle term, which has different meanings in the two premises (Dunbar 2001; Fearnside & Holter 1959; Hamblin 1970; Kroeger 2005, § 3.1; Quine 1960, §§ 27-31). If the middle term assumes a different meaning in

---

2To elucidate this classification of the arguments, some authors distinguish between deductive and inductive standards. If, from the truth of the premises, the conclusion necessarily follows, the argument is said to be **deductive**. Conversely, if the truth of the premises does not necessarily establish the truth of the conclusion, but nonetheless their truth provides good reasons to believe the conclusion to be true, then the argument is **inductive** (Govier 1980).

3 We say that a statement is plausible if it seems worthy of approval or probable.
each premise, then a syllogism, *de facto*, contains a fourth, hidden term, that causes the fallacy. Here is a simple example:

\[
(P_1) \text{Mick Jagger loves rock} \\
(P_2) \text{Rock is solid mineral material} \\
\text{hence} \\
(C) \text{Mick Jagger loves solid mineral material}
\]

Since *formal fallacies* are those arguments rendered invalid by a flaw in their logical structure (see, e.g., Hamblin 1970; Walton 1996), it is perhaps apt to quote Keynes (1887, p. 176): «The fallacy resulting from the ambiguity of one of the terms of a syllogism is a case of quaternio terminorum, i.e. a fallacy of four terms. This fallacy was often traditionally called the fallacy of ambiguous middle, and classified as a formal fallacy. But Schiller (1912, p. 365) pointed out that the fault is not a formal fallacy, in the sense of being simply an invalid syllogism, and is better classified as being a case of the fallacy of equivocation». Indeed, a quaternio terminorum falls in the set of formal fallacies because it really involves four terms and this does not really depend on the eventual ambiguity of any term involved. On the other hand, however, the case of quaternio terminorum is rather of a special nature. In fact, consider the following example from Mellone (1914):

\[
(P_1) \text{All metals are elements} \\
(P_2) \text{Brass is a metal} \\
\text{Hence} \\
(C) \text{Brass is an element.}
\]

where «the premises have no link of connection, and contain four different terms between them. Such mistakes are possible because of the ambiguity of language. If any term is used ambiguously, it is really two terms; hence the syllogism containing it has at least four terms, and is not a true syllogism at all, though at first sight it may appear to be one [...] using the middle term metal in two different senses, in one of which it means the pure simple substances known to chemists as metals, and in the other a mixture of metals commonly called metal in the arts, but known to chemists by the name alloy» (Mellone 1914, p. 166). If one follows Mellone’s interpretation, quaternio terminorum is also undoubtedly a fallacy of ambiguity or equivocation. Through our experimental results we will demonstrate that at least the perception of this type of fallacy does not only depend on its logical form but also on the (nature of the) ambiguity of the middle term.

1.1.2 Lexical pragmatics

The most common forms of lexical ambiguity are *homonymy* and *polysemy*. Homonymy refers to terms with different literal meanings bearing no semantic relation. Polysemy, instead, refers
to terms with literal meanings conveying semantic overlaps (Lyons 1977; Taylor 2003). An example of homonymy is the term “bank”, which may refer to a financial institution in “John went to the bank to open a savings account”, and to a riverside in “Plato and Socrates had a picnic on the bank”. An example of polysemy is the term “letter”, which means both “symbol of the alphabet” and “written communication”. In the case of homonymy, the meanings of the different occurrences do not share any property, while in case of polysemy some properties are shared because of the semantic overlap between the two meanings. This overlap is induced by different phenomena, among which are metaphors. Metaphors have been not only considered in connection with polysemy in cognitive semantics («the conceptual metaphor explains the systematicity of the polysemy», Lakoff & Johnson 1980, p. 248), but also as one of the most important way of creating new meanings (Bartsch 2002).

Metaphors are, at a linguistic level, words with multiple meanings. Take the word “grasp”, which can mean: “to hold on”, but also “to apprehend”, “to understand” and “to grip”. All these meanings are lexicalized (i.e. stored in a lexicon); we do not perceive them as “pregnant metaphorical uses” (Black 1993, p. 25). They are rather conventional uses, that scholars call “lexicalized metaphors” or “dead metaphors”. They form part of our conceptual maps and we find them in dictionaries. A “dead” metaphor is a lexical item with a conventional meaning different from the original (or some previous meaning in the chain of semantic change). Therefore, there is no need to consult the original meaning in order to understand it. For example the expression “falling in love” represents a dead metaphor and the occurrence of “to fall” has a conventional meaning: we do not need to know the original literal meaning to understand the sentence (Gola 2007).

“Live metaphors” are instead new and signify a creative use of language, neither to be found in common usage not already classified in dictionaries. Therefore live metaphors involve a creative, on-the-fly, comprehension process. However, in the conceptual theory of metaphor (Lakoff & Johnson 1980), live metaphors are also supposed to be as much alive as the conventional and vital conceptual metaphors in which they are considered to be grounded in. For example, the conceptual metaphor LIFE IS A JOURNEY gives rise to several conventional meanings: as “I do not know which path to take”, but also unconventional, poetic utterances like the verses of Robert Frost’s poetry “The Road Not Taken” (1920):

Two roads diverged in a wood, and I...
I took the one less traveled by,
And that has made all the difference.

Lakoff and Turner (1989) have shown many similar examples, maintaining that «great poets can speak to us because they use the modes of thought we all possess», and that «to understand the nature and value of poetic creativity requires us to understand the ordinary ways we think» (1989, pp. xi-xii).

The distinction between dead and live metaphors is controversial. Lakoff and Johnson (1980), for example, maintain that metaphors do not cease to influence our thought and action just because we are not aware of them. This issue is at the core of the current debate on linguistic and conceptual theories of metaphor, as Müller (2008) highlights, claiming that there is a grey area of metaphors, which range from dead to live metaphors. Occasionally a
conventional metaphor may be “delexicalized” (Pawelec 2006), or a dead metaphor can be “revitalised” (Alm-Arvius 2006), eliciting a creative comprehension process. However, as long as the context does not require it, dead metaphors go unnoticed, as in the case of a literal polysemy.

2. Design and predictions

The study was designed to investigate whether, and to what extent, lexical ambiguity (such as homonymy, polysemy and metaphor) influences the detection of a quaternio terminorum, which is based on the intrinsic ambiguity of the middle term. The experimental design comprised four groups of ambiguous middle terms, classified as follows: homonymy (H), polysemy (P), dead (lexicalized) metaphor (DM), live metaphor (LM). From now on, with H, P, DM, and LM, we shall denote the classes of arguments featuring homonymous terms, polysemous terms, dead (lexicalized) metaphors, and live metaphors, respectively. We planned a set of arguments with H, P, DM, LM middle terms, having the structure of a quaternio terminorum. We proposed arguments with the following structure:

Premise 1: \( x \) verb \( y \);
Premise 2: \( y \) verb \( z \);
Conclusion : \( x \) verb \( z \).

In order to spell out the perception of the strength of an argument, we asked participants to evaluate whether the conclusion follows from the premises.

The two guiding-questions of our study were the following:

1. To what extent does lexical ambiguity influence the detection of a quaternio terminorum?
2. Which case of ambiguity affects its persuasiveness?

As to the first question, our prediction was that quaternio terminorum identification should mainly depend on the nature of the middle term, and therefore on the degree of partial semantic overlapping between the different readings of a middle term (degree of shared semantic properties). In particular, we expected that arguments featuring homonymous words (e.g. “bank”) as middle terms would be more easily recognized as fallacious than arguments featuring polysemous terms (e.g. “letter”) or dead metaphors (e.g. “star”) as middle terms. This might be due to the fact that different meanings of a homonymous term would be clearly divergent, whilst the meanings of a polysemous or metaphorical term partly overlap (Ervas & Ledda 2014; Gentner, Ratterman, Forbus 1993; Gick & Holyoak 1983). As to the second question, we expected arguments featuring polysemous or metaphorical terms to be more persuasive. As an effect of them being lexicalized, premises with dead metaphors may appear true even if they are literally false. Consequently, we expected quaternio terminorum with dead metaphors as middle terms to be most persuasive. We expected instead that premises with live metaphor were easily recognized as literally false, because the figurative meaning of a live metaphor clearly differs from its literal meaning, thus making the ambiguity easily detectable.
Empirical evidence was needed to understand whether and to what extent participants’ intuitions on the strength of an argument vary according to the category to which the middle term belongs. To avoid any bias in participants’ understanding of the overall argument, we did not explicitly teach them how to distinguish between a strong and a weak argument. To understand whether participants would have been able to distinguish between a strong and a weak argument, we planned a set of distractors having the structure of a standard quaternio terminorum featuring no ambiguous terms. The set of distractors comprised clearly strong arguments with true premises/true conclusion and clearly weak arguments with true premises/false conclusion. The role of distractors was twofold:

1. reflecting participants’ understanding and commitment to the task they were assigned;
2. emphasizing participants’ capacity to distinguish between a strong and a weak argument, without any explicit instruction.

Finally, we accepted only those participants who were able to distinguish between strong and weak arguments (acceptance threshold: more than 90% of correct answers).

We also needed to overrule the possibility of participants finding alternative strategies to understand whether the conclusion follows from the premises. In the experimental design we indeed realized that a unique set of standard quaternio terminorum with true premises and false conclusions could have been misleading in participants’ evaluation of the strength/weakness of an argument for at least two reasons. First, since the middle term possesses different meanings in the two premises, we had to make sure that participants could distinguish between middle terms needing disambiguation in the two premises, and middle terms unambiguously used. Therefore, we also prepared a set of strong arguments with true premises, true conclusions and a middle term used with the same meaning in both premises. Arguments in this set could appear fallacious because of the quaternio terminorum form, but they are indeed strong. In this way we could check participants’ ability to distinguish between weak arguments (standard fallacies of the four terms) and strong arguments with unambiguously used middle terms. Second, and more interestingly, we had to make sure that participants did not consider an argument weak just because of its false conclusion. Indeed, a patently false conclusion could lead participants to judge an argument weak without any real understanding. To this effect we prepared a set of weak arguments whose middle term had different meanings in the premises, as in standard quaternio terminorum case, but whose conclusions were plausible. This was to avoid the effect a patently false conclusion could have on participants’ intuitions.

To summarize the experimental design: we planned to ask the participants to judge the strength of the following sets of arguments (argument structure condition) combined with H, P, DM, LM middle terms (middle term condition):

1) a set of standard quaternio terminorum (with 6 X H, P, DM, LM middle terms) with true premises/false conclusion;
2) a set of strong arguments (with 6 X H, P, DM, LM middle terms) with true premises/true conclusion;
3) a set of quaternio terminorum with plausible conclusion (with 6 X H, P, DM, LM middle terms) with true premises/plausible conclusion.
Since their middle terms needed no disambiguation, we hypothesized that the strength of the arguments belonging to the second set would be easily detected by participants. Moreover, because participants would had been forced to assess the conclusion by considering whether it actually follows from the premises, we expected that the weakness of the arguments in the third set would be spotted with greater difficulty than in standard quaternio terminorum case.

3. Method

We tested premises and conclusions of three sets of arguments in three norming studies. The method included a primary selection of a set of middle terms (=206 nouns), that could possibly be used to form middle terms belonging to the four categories H, P, DM, LM, according to their number of letters and frequency in the GRADIT (De Mauro 2000). In the first norming study, we tested both emotional (positive and negative) meaning and familiarity of the selected terms using a 1 (very negative/very unfamiliar) to 5 (very positive/very familiar) rating scale. We eliminated the terms with definite emotional meanings (“positive meaning” mean > 4 and “negative meaning” mean < 2) and insufficient familiarity (mean < 3).

We used unambiguous terms to build live metaphors and then devised arguments on the basis of the chosen middle terms. In the second norming study, we tested the premises of the arguments separately in order to make sure that participants attribute either the same meaning to the middle terms in case of strong arguments or different meanings in case of quaternio terminorum. In the third norming study, we tested the premises and conclusions separately in order to understand whether they were perceived as either true or false. The results of norming studies have shown that the majority of sentences featuring a dead metaphor (83%) are perceived as true, even if they are literally false, whilst the majority of sentences featuring a live metaphor (79%) are perceived as false. The results of the norming studies induced us to exclude, from the experimental design, LM middle terms: being perceived as false renders them incomparable with the other categories of middle terms. Therefore, the pilot study had a 3X3 experimental design: 3 argument structure conditions X 3 (H, P, DM) middle term conditions.

3.1 Participants

All participants were native Italian speakers, and had normal/corrected vision. They signed informed consent agreements to volunteer for this study, which complies with APA ethical standards. The norming studies involved 209 participants (balanced across age and gender). All of these participants were undergraduate students at the University of Cagliari (Faculty of Education, Languages and Engineering) and did not participate in the pilot study. The pilot study involved 40 participants (balanced across age and gender), who were undergraduate students at the Faculty of Humanities, University of Cagliari, attending the first year BA Program in Communication Science. We made sure that participants were naïve, i.e. that they had not previously attended any logic course.
3.2 Materials

The printed materials of the pilot study consisted of a set of 54 arguments and 50 distractors in Italian. The 54 arguments comprise 18 arguments with quaternio terminorum’s structure for each middle term condition: homonymy (H), polysemy (P) and dead (lexicalized) metaphor (DM). For each group of ambiguous middle terms, 6 arguments were standard quaternio terminorum (with true premises/false conclusion), 6 arguments were strong arguments (with true premises/true conclusion) and 6 arguments were quaternio terminorum with plausible conclusion (with true premises/plausible conclusion), as in the examples in English in Table 1.

Table 1. Examples in English for each argument structure condition combined with H, P, DM middle term conditions

<table>
<thead>
<tr>
<th>Homonymous middle terms (H)</th>
<th>Example of standard quaternio terminorum</th>
<th>Example of strong arguments</th>
<th>Example of quaternio terminorum with plausible conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>[P1] A large dancing party is a ball; [P2] A ball is a sphere; [C] A large dancing party is a sphere.</td>
<td>[P1] Boston marathon is a race; [P2] A race is a competition; [C] Boston marathon is a competition.</td>
<td>[P1]; A financial building is a bank; [P2] A bank is at river’s edge; [C] A financial building is at river’s edge.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Polysemous middle terms (P)</th>
<th>Example of standard quaternio terminorum</th>
<th>Example of strong arguments</th>
<th>Example of quaternio terminorum with plausible conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>[P1] Liverpool Street is a lane; [P2] A lane is a part of a swimming pool; [C] Liverpool Street is a part of a swimming pool.</td>
<td>[P1] A border is a limit; [P2] A limit is the end of an area; [C] A border is the end of an area.</td>
<td>[P1]; An alphabet symbol is a letter; [P2] A letter is a written communication; [C] An alphabet symbol is a written communication.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dead (lexicalized) metaphors as middle terms (DM)</th>
<th>Example of standard quaternio terminorum</th>
<th>Example of strong arguments</th>
<th>Example of quaternio terminorum with plausible conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>[P1] Jimi Hendrix is a legend; [P2] A legend is a traditional story; [C] Jimi Hendrix is a traditional story.</td>
<td>[P1] Clooney is a star; [P2] A star is a famous actor; [C] Clooney is a famous actor.</td>
<td>[P1]; A hot place is an oven; [P2] An oven is part of a kitchen; [C] A hot place is part of a kitchen.</td>
<td></td>
</tr>
</tbody>
</table>

The 50 distractors comprise 25 clearly strong and 25 clearly weak arguments, as in the examples in English in Table 2.
Table 2. Examples in English of distractors (clearly strong arguments and clearly weak arguments)

<table>
<thead>
<tr>
<th>Examples of clearly strong argument</th>
<th>Example of clearly weak argument</th>
</tr>
</thead>
<tbody>
<tr>
<td>[P1] Francesca is a woman;</td>
<td>[P1] Antonio hates wind;</td>
</tr>
<tr>
<td>[P2] A woman is a female;</td>
<td>[P2] Wind moves the mills;</td>
</tr>
<tr>
<td>[C] Francesca is a female.</td>
<td>[C] Antonio moves the mills.</td>
</tr>
<tr>
<td>[P1] Elisabetta has a doll;</td>
<td>[P1] Giuseppe eats an apple;</td>
</tr>
<tr>
<td>[P2] A doll is a toy;</td>
<td>[P2] An apple is a fruit;</td>
</tr>
<tr>
<td>[C] Elisabetta has a toy.</td>
<td>[C] Giuseppe is a fruit.</td>
</tr>
</tbody>
</table>

3.3 Procedure

In the preparation for the printed test, we randomized the arguments and the distractors. We presented to the participants a paper/pencil test with two columns: a column for the materials (both arguments and distractors) and a column with a “YES” and a “NO” box for each argument/distractor. Participants were asked to judge, for each argument/distractor, whether the conclusion [C] follows from the premises [P1] and [P2], by checking the “YES” box if they thought that [C] did follow from [P1] and [P2], or the “NO” box otherwise. We registered the “YES”/“NO” answers. The overall test lasted for a maximum of 25 minutes.

4. Results

We performed a t-test to determine the statistical significance. As regards distractors, in more than 90% of cases (see table 3), all participants answered “Yes, the conclusion follows from the premises” in presence of a clearly strong argument (mean difference = 37.92; \( p < 0.001 \)), and “No, the conclusion does not follow from the premises” in case of a clearly weak argument.

Table 3. Mean of answers for clearly strong and clearly weak arguments as distractors
Results show a significant difference of performance between the “strong arguments” (TP/TC) condition and the set of “standard quaternio terminorum” (TP/FC) condition. This significant difference is due to greater number of correct responses (“YES”) to the “strong arguments” condition compared to the number of correct responses (“NO”) to the “standard quaternio terminorum” condition for homonymous middle term (mean difference = 11.67; \( p < 0.001 \)), polysemous middle term (mean difference = 12.5; \( p < 0.001 \)) and dead metaphors middle term (mean difference = 12.17; \( p < 0.001 \)) conditions (see table 4). In both cases of strong arguments and standard quaternio terminorum no significant difference was found among homonymy, polysemy and dead metaphors conditions.

Table 4. Mean of answers for strong (TP/TC) and weak (TP/FC – standard quaternio terminorum) arguments with H, P, DM middle terms

![Bar chart showing mean of answers for strong (TP/TC) and weak (TP/FC – standard quaternio terminorum) arguments with H, P, DM middle terms.]

Legenda for table 4:
TP = true premises;
TC = true conclusion;
FC = false conclusion;
PC = plausible conclusion.

In case of quaternio terminorum with plausible conclusion (TP/PC) condition, no significant difference was found between incorrect answers (“YES”) and correct answers (“NO”) in both homonymy (mean difference = -5.66; \( p = 0.25 \)) and polysemy condition (mean difference = 9; \( p = 0.13 \)). A significant difference (see Table 5) was observed between incorrect answers (“YES”) and correct answers (“NO”) in dead metaphors condition (mean difference = 17.34; \( p = 0.015 \)). Furthermore, in the TP/PC condition, a certain trend towards significance was found while comparing the number of correct answers (“NO”) in the homonymy condition and the number of correct answers (“NO”) in the polysemy condition (mean difference = 7.33; \( p = 0.08 \)), whilst a
significant difference was found in the number of correct answers ("NO") within the homonymy condition and the dead metaphors condition (mean difference = 11.51; \( p = 0.005 \)). Instead no significant difference was found between the number of correct answers ("NO") in the polysemy condition and the dead metaphor condition (mean difference = 4.17; \( p = 0.29 \)).

Table 5. TP/PC cases across categories

![Graph showing TP/PC cases across categories (Homonymy, Polysemy, Dead Metaphors)](image)

5. Discussion

The study aimed to investigate whether lexical ambiguity influences elementary inferences in argumentation. The results from the pilot study suggests that, in general, participants can evaluate correctly the strength of an argument not only in the presence of non-ambiguous middle terms (as in case of distractors, see Table 3), but also in the presence of potentially ambiguous middle terms with the same meaning in both premises (as in the case of “strong arguments”, see table 4). The majority of participants are indeed able to identify strong arguments even in presence of potentially ambiguous middle terms. They are also able to single out standard quaternio terminorum with true premises and false conclusion. It should be observed, however, that compared to the identification of strong arguments with potentially ambiguous middle terms their performance is slightly inferior.

On the one hand, the results reveal an effect of the logical structure: if compared with strong arguments with ambiguous middle terms, used with the same meaning in both premises, standard quaternio terminorum are much harder to detect.(see § 5.1). On the other hand compared to standard quaternio terminorum the results report an effect of lexical ambiguity in the detection of fallacious arguments: quaternio terminorum, with plausible conclusion, are harder to detect (see § 5.2).
5.1 The detection of lexical ambiguity fallacy

Why arguments with univocally used middle terms were easily detectable compared to fallacious arguments? In our view, arguments with no lexical ambiguity are easier to detect since they involve no disambiguation process and require just three terms in the logical structure. On the contrary, a quaternio terminorum requires a process of disambiguation to realize the existence of four terms in the logical structure: participants should suppress one of the two literal meanings of the middle term, which means something in the first premise and something else in the second. Disambiguating a homonymous middle term would require suppressing one of its two literal meanings, namely the irrelevant one (Gernsbacher 1990). For instance, disambiguating the homonymous word “bank” would require selecting one of its two meanings, i.e. financial institution or river side (Gernsbacher 1990; Gernsbacher & Faust 1991). Therefore processing the lexical form “bank” requires the activation of two different and unrelated lexical entries, and the suppression of the irrelevant one. In the case of homonymy, the selection of the relevant meaning is performed by default on the basis of the so-called "pre-semantic or narrow context" (Perry 1997, 2001). In the case of polysemy, the selection of the relevant meaning involves a broader context (Bach 2012; Carston 2002; Perry 1997, 2001; Recanati 2004). Contrary to homonyms, where completely independent meanings are involved, polysemous words exhibit a list of possible (related) meanings that might be selected.

Appealing to a “unified approach” to literal and non-literal usages of language, a process of modulation has been proposed by Carston (2002) and Recanati (2004, 2010), among others, to explain the process of selecting the relevant meaning in polysemy, and metaphors. For polysemy, the process of modulation can be a process of lexical narrowing or broadening. We speak of lexical narrowing when the concept expressed by the usage of a term conveys a more restricted interpretation than the linguistically-encoded concept. For instance, in the sentence “I cut the grass”, we pick up a more specific concept “cut” than the conventional concept encoded by the polysemous term “cut”. This “ad hoc” concept is relative to the sentential context and differs from the “ad hoc” concept we pick up in the sentence “I cut my hair”. We speak of lexical broadening when the concept expressed by the usage of a term conveys a more general interpretation than the linguistically-encoded concept. For instance, in the sentence “It’s boiling outside”, the linguistically-encoded concept “boiling” is contextually adjusted in a pragmatic process of modulation resulting in the ad hoc concept “very hot” (Falkum 2011). A metaphorical extension is a type of broadening, where some properties of the linguistically-encoded concept are selected to grasp the intended ad hoc concept relevant to the context. For instance, to understand the sentence “Jimi Hendrix is a legend”, we merely select the properties of the linguistically-encoded concept “legend” required to grasp the intended ad hoc concept, i.e. to be fabulous, well-known, or extra-ordinary in a certain field. Other irrelevant properties of the linguistically-encoded concept “legend” (for instance, to be a traditional story) are suppressed (Glucksberg, Newsome & Goldvarg 2001; Rubio Fernandez 2007). In this perspective, metaphors are explained as a local, on-line pragmatic adjustment of the encoded lexical meaning that results in an ad hoc concept which conveys the figurative meaning.
5.2 Lexical ambiguity in quaternio terminorum with plausible conclusion

As expected, the most interesting results come from the set of quaternio terminorum with true premises and plausible conclusion (see Table 5), which forces participants to evaluate the whole argument and avoid a separate reading of the single conclusion, as in case of standard quaternio terminorum. In the case of fallacious arguments, with a plausible conclusion, the results of the pilot study suggest some difficulties in recognizing the fallacy in case of polysemous middle term condition, even though there is no meaningful difference but merely a trend towards significance is registered compared to the homonymous middle term condition ($p = 0.08$). This result could be motivated by referring to the modulation process required to disambiguate a middle term. On the one hand, a homonymous middle term can be disambiguated by a default selection of one of its clearly distinct meanings. On the other, a polysemous middle term may present a semantic overlapping area among its meanings which makes it more difficult and demanding to detect the meanings used in the two premises. In quaternio terminorum with a plausible condition, this process will be more effective since, in order to understand whether the conclusion really follows from the premises, the participants are forced to disambiguate the meanings of the middle term.

In case the middle term is a dead metaphor, we observe a very significant difference between the number of correct answers given in the dead metaphors middle term condition and in the homonymous middle term condition ($p = 0.005$). Therefore, compared to homonymous terms, dead metaphors seem to have a strong influence on the evaluation of the overall argument, and hence prompt an alternative style of reasoning. On the contrary, dead metaphors middle terms condition turns out to be not significantly different from polysemous middle terms condition ($p = 0.29$). This is probably because they share a similar process of modulation, and therefore a similar mechanism of suppression of irrelevant information, as previously argued (see § 5.1). In any case, a significant difference ($p = 0.015$) between participants missing and recognizing the fallacy can be observed just in dead metaphors middle term condition, and not in homonymous and polysemous middle term conditions. This seems to suggest that, in the case of dead metaphors as middle terms, participants are more prone to judge a fallacious argument as strong. The results therefore suggest that dead metaphors do alter participants’ perception of the strength of an argument and influence their intuitions on the logical connection of the whole argument, thus making them to consider it strong.

These results can be explained by distinguishing between the truth conditions of a literal sentence, and the intuitive truth conditions assigned by a speaker in specific contexts in which a sentence is used (Ervas & Ledda 2014). While we can assign truth conditions to premises where a literal (homonymous or polysemous) middle term occurs, we do not assign truth conditions to premises containing a non-literal, metaphorical middle term. In case of metaphorical middle terms we perceive premises as true, thus assigning them intuitive truth conditions, even though they are literally false. According to Contextualism and Relevance Theory (Carston 2002; Recanati 2004, 2010; Sperber & Wilson 1986/1995), understanding a statement is knowing the concrete circumstances of its truth. In this view, the pragmatic process involved in dead metaphor comprehension takes the encoded concept and generates an ad hoc concept in the proposition the speaker intends to communicate, i.e. a proposition corresponding to the intuitive truth-conditions assigned by the speakers (Carston 2002). The “falsehood” of a
metaphor is indeed a “myth”, and an attempt to judge the metaphor under some sort of truth conditions, the literal ones, that cannot explain the very nature of the metaphor itself (Clark 1994). Therefore there would be no literal meaning in people’s intuitions: when participants read a premise where a metaphor occurs, they assign intuitive truth conditions to the sentence, thus immediately modulating the metaphorical term and considering the sentence in which it occurs at least plausible, if not true. The contribution of a metaphor to the overall truth-conditions of a sentence is then its intuitive truth-conditions, which respect speaker’s semantic intuitions, thus diverging from the predictions of classical argumentation. This could be the reason why speakers judged premises featuring a dead metaphor as true, thus compromising the evaluation of the strength of the overall argument, when they are forced to judge whether the conclusion follows from the premises.

However, on the non-contextualist side, it could be claimed that dead metaphors are just perceived as true because they are lexicalized, as in the case of literal meanings, such as in polysemy (Stern 2006; Szabo 2012). Proper, live metaphors would still be perceived as false, as the classical view predicts (Grice 1989; Katz 1972; Katz & Fodor 1963; Searle 1985). This is one of the main reasons why we could not compare live metaphors with other cases of lexical ambiguities in this experimental study. We could anyway imagine that live metaphors also might be perceived to be true, when a broader context is presented. Indeed, we could suppose that the identification of fallacious arguments, with live metaphors as middle terms, would have been significantly different. The experimental literature has shown that the interpretation process of novel metaphors diverges from that of conventional metaphors (Blasko & Connine 1993; Thibodeau & Durgin 2008). Indeed, live metaphor comprehension involves elaborate pragmatic processes – for instance iconic representations of concepts or imagery (Carston 2010; Indurkhya 2007). Because of their unfamiliarity, a wider context is required to understand them. The contextual information in a sentence would be too narrow to produce the typical imagistic effect live metaphors possess (Lai, Curran & Menn 2009). If Aristotelian standards of syllogisms are respected, in argumentative contexts such as those represented by the concatenation of premises/conclusion in a quaternio terminorum, live metaphors have a very narrow context to be interpreted in and would be easily recognized as false. From our point of view, there should be a strict link between the evaluative truth conditions of the premises and the overall identification of the strength of the whole argument. In a narrow context, dead metaphors are perceived as true even though they are literally false. The vast knowledge presupposed by an everyday use of our mother language is sufficient to recognize the relevant properties carried by the conventional metaphor, a broader context is not necessarily required (Glucksberg & Estes 2000). The case of lexicalized metaphors is indeed very interesting because, as the experimental literature shows, they are processed as fast as literal meanings (Giora 2003) and participants show some difficulty in rejecting them as literally false (Glucksberg 2003). This might be the main reason why “common” dead metaphors make fallacious arguments persuasive. Therefore, it is plausible that difficulties in attributing truth conditions to premises featuring metaphoric ambiguity influence the detection of the strength of the whole argument, as is the case with quaternio terminorum.
6. Conclusion

The results of our study show that it is more difficult to detect a meaning ambiguity fallacy in comparison with a strong argument. In case of lexical ambiguities, we suppose that the reason behind this phenomenon is that the disambiguation processes, required to single out the different meanings of the middle term in the premises, could complicate the task of identifying the fallacy. Indeed, arguments of the type of quaternio terminorum may indeed contain either lexically ambiguous (homonymous or polysemous) or metaphorical middle terms, which ask for additional attentional resources. As we mentioned, our experimental findings indicate that at least the perception of a quaternio terminorum relies very much on the (nature of) ambiguity of the middle term.

Especially in case of dead metaphors, our results seem to suggest that participants have difficulties in detecting the fallacy of the four terms. Interestingly, this is the case when they are asked to verify the connection between premises and the conclusion of an argument. We argued that this could be due to the fact that participants assigned intuitive truth conditions to premises where a dead metaphor occurs, thus considering true, premises that were literally false. Determining either truth or falsity of premises with lexical ambiguities represents a major step towards the identification of the type of an argument. This process seems to influence the detection of a quaternio terminorum. It is common to judge sentences featuring dead metaphors as true, relying on naif intuitions. This phenomenon affects the global understanding of an argument, and often leads participants to claim that a fallacy is a strong argument.

Moreover, for quaternio terminorum with a plausible conclusion, fallacies with polysemous middle terms are quite difficult to detect in comparison to fallacies involving homonymous middle terms. An explanation for this phenomenon can be found in the modulation process required, in case of polysemy and dead metaphor, to suppress the irrelevant meaning. The process of modulation would operate on a list of overlapping meanings, while the process of homonymous terms disambiguation would operate by default on completely different meanings. Further research is required to understand better the case of polysemy in comparison to the case of both metaphor and homonymy. Indeed, the suppression mechanism seems to operate faster in the resolution of homonymy than in polysemy (and dead metaphors) interpretation (Gernsbacher & Faust 1991; Gernsbacher et al. 2001; Rubio Fernandez 2007). Participants’ reaction times in the evaluation of arguments, with homonymous and polysemous middle terms, would then be precious to understand better the influence different forms of lexical ambiguities may have in the detection of fallacies.

Finally, the experimental study excluded live metaphors as middle terms. In fact, in the norming studies participants considered premises, where a live metaphor occurs, patently false. This made them incomparable to the other group of arguments with true premises. Our hypothesis is that the detection of an argument, whose middle term is a live metaphor, depends on the broadness of the context provided. This could explain why, within the narrow context involved in the premises, participants interpreted live metaphors as literally false. In any event, because of their evident falsity, they should not be an issue for the detection of the strength of the overall argument. In contrast, a narrower context would be sufficient for the case of dead metaphors, because of their high familiarity and frequency. For the same reason, dead metaphors can be retained as true, thus causing an incorrect evaluation of the overall argument.
In order to compare the effects dead and live metaphors have in the detection of fallacies, future research should consider wider contexts in the premises of the arguments.

Acknowledgements

This work is the outcome of a collaborative effort. However, for the specific concerns of Italian academy, Francesca Ervas is responsible for sections 4-5, Elisabetta Gola for the section 1.1.2; Antonio Ledda for the sections 2-3 and 6, Giuseppe Sergioli for the sections 1, 1.1 and 1.1.1. We warmly thank Bipin Indurkhya, Francesco Paoli and Maria Grazia Rossi for the stimulating discussions. We also thank the two anonymous reviewers for their constructive comments, which helped us to improve the manuscript. Francesca Ervas gratefully acknowledge the support of the Sardinia Regional Government for the financial support (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2007-2013 – Axis IV Human Resources, Objective I.3, Line of Activity I.3.1). Antonio Ledda and Giuseppe Sergioli gratefully acknowledge the support of the Italian Ministry of Scientific Research within the FIRB project “Structures and dynamics of knowledge and cognition”, Cagliari-F21J12000140001. Elisabetta Gola gratefully acknowledges the support of the Sardinia Regional Government (L. 7/2007) for the financial support within the project “Argomentazione e Metafora. Effetti della comunicazione persuasiva nel territorio sardo”.

References


Keynes, J.N. 1897: *Studies and Exercises in Formal Logic*, London, Macmillan and Co..


