Running title: Regular and novel metonymy acceptability

Regular and novel metonymy in native Korean, Spanish and English: Experimental evidence for various acceptability

Roumyana Slabakova[1,2]

Jennifer Cabrelli Amaro[3]

Sang Kyun Kang[2]

[1]University of Southampton, [2]University of Iowa, [3]U. S. Air Force Academy

Address for correspondence:

Roumyana Slabakova
Department of Modern Languages
Building 65
Avenue Campus
University of Southampton
Highfield
Southampton
SO17 1BF
United Kingdom

R.Slabakova@soton.ac.uk

Regular and novel metonymy in native Korean, Spanish and English: Experimental evidence for various acceptability

Abstract

This article presents results of two off-line comprehension tasks, investigating the acceptability of novel and regular metonymy by speakers of English, Korean, and Spanish. We are interested in uncovering regular-novel metonymy computation discrepancies, and whether they are treated differently in the different languages. The distinction between novel and regular metonymy is discussed by the existing theoretical treatments of metonymy as well as in psycholinguistic research. The findings of this study constitute further experimental support for the psychological reality of this distinction. In addition, it is demonstrated that the speakers of the three languages treat novel and regular metonymy differently. Significant findings are the acceptability of novel metonymy in Korean and the relative lack of conventionalization effect for regular metonymy in Korean and Spanish. We conclude that current theoretical approaches to metonymy should focus more on cross-linguistic differences and that further language comparisons are warranted and needed, in comprehension as well as in processing.

Keywords: metonymy comprehension, regular and novel metonymy, Korean, Spanish, English

1. Introduction

The second presidential debate of the 2012 election season between President Barack Obama and Governor Mitt Romney was held on October 16, 2012. Governor Romney said the following:

> "And I—and I went to my staff, and I said, 'How come all the people for these jobs are—are all men.' They said: 'Well, these are the people that have the qualifications.' And I said: 'Well, gosh, can't we—can't we find some—some women that are also qualified?' And—and so we—we took a concerted effort to go out and find women who had backgrounds that could be qualified to become members of our cabinet. I went to a number of women's groups and said: 'Can you help us find folks,' and they brought us **whole binders full of women**."

Governor Romney used the perfectly legitimate linguistic process of transfer of reference, or metonymy, where 'binders full of women' actually stands for 'binders full of women's CVs, or dossiers.' However, the reaction of the public showed that this meaning transfer, completely legitimate in the language but novel, ad-hoc, not-heard-before, may not be that readily available to the comprehending public. The phrase then took on a life of its own: It started a meme on the internet, it gave rise to countless tweets, Tumblr postings, it even got its own page on Amazon.com. Challengers of the phrase thought it was insensitive and insisted on interpreting it literally, as the following Tumblr picture illustrates.

The phrase and the subsequent outcry illustrate an interesting linguistic fact. Metonymy is a well-established mental process, whereby the mention of some entity (activity, person, thing, time period, etc.) is interpreted *to stand for* a related entity. Typical substitutions of this sort include an author for his/her work (e.g., 'Have you read the latest Alice Munro?'), a capital name for the country government (e.g., 'Paris is in a huff'), and a place name for a typical product (e.g., 'We drank Côtes-du-Rhône last night'). These metonymies are widely conventionalized and many speakers produce and comprehend them effortlessly and freely. In this article, we will dub them instances of 'regular metonymy' because they reflect recurrent, entrenched conceptual mappings such as PART FOR WHOLE, CAUSE FOR EFFECT, PERSON FOR ROLE, PLACE FOR EVENT, etc.[1] However, other metonymies, such as the 'binders full of women' illustrated above, are not widely conventionalized although they use the same mental processes. Many of them are

[1] The literature on cognitive linguistics usually presents metonymic cognitive patterns in small caps, as we have done here.

produced and comprehended online; that is, they are novel and their interpretation depends on the concrete context of the utterance together with linguistic and pragmatic principles of interpretation. Such metonymies constitute the wider transfer of reference calculation, whose classical example is Nunberg's (1979) sentence 'The ham sandwich is getting restless.' Spoken between two waitresses in a diner to identify a customer who has ordered a ham sandwich, the utterance makes perfect sense, although it may not be readily comprehensible in another context. We will label such utterances 'novel metonymy,' being mindful of the fact that regular and novel metonymy are not mutually exclusive, but rather two opposites on a cline of metonymy conventionalization.

The linguistic literature on metonymy is extensive, starting with Nunberg's radical pragmatics account in 1979. In recent decades, researchers have also been rightfully concerned with the question of whether cognitive or linguistic theories provide psychologically real accounts of how people construct metonymic, transferred meaning (Gibbs, 2009). However, the extensive cognitive, linguistic, and psycholinguistic literature is mostly based on the English language and predominantly cites examples of metonymy from the authors' linguistic experience. Corpus studies are rare (but see Brdar-Szabó, & Brdar 2003; Markert & Nissim, 2003 for some exceptions). In addition, as recently claimed by Gibbs (2009), "there are virtually no experimental studies that have explicitly attempted to find evidence on the role of conceptual metonymies in figurative meaning construction." (Gibbs, 2009: 33). In an attempt to address this gap in our knowledge, this article presents an experimental study on the comprehension of regular and productive metonymy by English, Spanish, and Korean native speakers. Cognitive linguistics (Croft, 2002/1993; Kövesces & Radden, 1998; Lakoff, 1987; Lakoff & Johnson, 1980; Panther & Thornburg 2007) claims that metonymy is a universally

available mechanism, although it is dependent on conceptual structures, cultural knowledge, and pragmatic routines. Our purpose was to establish whether speakers of these three languages engage in similar processes of figurative language comprehension and whether the relatively more regular as well as the more productive metonymic mappings are readily available in these languages. We compared acceptable metonymic patterns with unacceptable ones and with baseline sentences without metonymies. Our main findings reveal that there is a lot more speaker variation than the literature acknowledges or is aware of, and that regular metonymic shifts are rated much higher, hence comprehended more easily, than novel, productive shifts.

The rest of the article is structured as follows. Several major approaches to metonymic expressions will be discussed in Section 2. Investigations of the psychological reality of metonymic language interpretation should crucially contribute to the validation of linguistic analyses, therefore the findings of several recent neuro- and psycholinguistic studies will be discussed in Section 3. We will specify the predictions of the existing accounts of metonymy as to our judgment elicitation in Section 4. The experimental study is described in Section 5, while Section 6 offers discussion of the results and some conclusions.

2. Theoretical accounts and experimental evidence of metonymy computation

In this section, we briefly outline the way metonymy is treated within cognitive linguistics by the radical pragmatics approach, and within lexicalist theorizing with a natural language processing slant. Lakoff and Johnson (1980) is the seminal publication within cognitive linguistics, which proposed that metaphor and metonymy should be thought of as the fundamental mechanisms of language and thought. Especially in the last decade, the study of metonymy has been propelled into a dominant position in cognitive linguistic studies (see e.g., Benczes, Barcelona & Ruiz de

Mendoza, 2011; Croft, 2002/1993; Kövesces & Radden, 1998; Lakoff, 1987; Panther & Thornburg, 2003, 2007; Panther, Thornburg & Barcelona, 2009). The literature on metonymy within cognitive linguistics is vast, and we cannot do it justice here. What most cognitive linguists agree on presently is the fundamentally conceptual nature of metonymy, the fact that it is experientially grounded, and that it involves experientially and conceptually connected 'contiguous' elements (Barcelona, 2011: 48-9). In a recent attempt to achieve a consensus view, Barcelona (2011) argued that metonymy is a cognitive process involving an asymmetric mapping relationship, not a 'stand-for' relationship. This mapping consists of the activation of the target concept (e.g., the books written by Alice Munro) from the perspective of the source (e.g., Alice Munro, the writer). A useful way for describing various cases of metonymy within this theoretical approach is the continuum of metonymicity (Barcelona, 2003), based on displaying more or less prototypical metonymy properties. Although this approach to metonymy is grounded in describing cognitive shifts and ratiocination, it has not produced much experimental data on how speakers of a language process the various types of metonymy online.

One of the most important questions related to the production and comprehension of metonymy is why some metonymies are viable and attested while some others are not. To use the well-known 'ham sandwich' metonymy mentioned above, we do associate this designation with a customer who has ordered a ham sandwich in a diner, thus the customer is contextually related to the sandwich; at the same time, we do not normally use 'the laptop' to refer to a person, although laptops are very frequently associated with their users, people. In other words, descriptions of cognitive rules and processes should not overgenerate, allowing infelicitous metonymies into the grammar. Panther and Thornburg (2003), still within a cognitive linguistics approach, propose that the comprehension and generation of metonymy is guided by certain

'inferential principles,' or 'conceptual metonymies' such as PART-WHOLE, CAUSE-EFFECT, PERSON-ROLE, REPRESENTATION-REPRESENTED. In their understanding, these reasoning principles are conceptual devices not restricted to language but also used in other semiotic systems. If a substitution is attempted that is outside, or not supported by these ready-made and accessible inferential principles, it will not go through, or will not be comprehended adequately (think the 'binders full of women' phrase discussed in the introduction.)

The radical pragmatic approach to metonymy (Fauconnier, 1985; Nunberg, 1995; Papafragou, 1996) has been among the first to tackle the question of why some shifts of meaning are attested and comprehensible while others do not go through. According to this approach, a change of sense is licensed when there is an important relation between the two meanings, based on interlocutors' knowledge of the world and context situation (Nunberg, 1995). For example, 'Have you read Alice Munro?' is acceptable because there is a noteworthy link between the work and the author. Radical pragmatic theories also rely on regular polysemy rules such as PRODUCER-PRODUCT and others, because the relationship between the producer/author and the work is noteworthy. A similar proposal is put forward by Papafragou (1996), who singles out salience in the discourse situation as the important relationship supporting metonymy construal. In addition, Nunberg (1995) proposes that there are language-specific polysemy conventions, which would presumably make the metonymies sanctioned by them easier to process than more irregular, rarer uses.

Within generative linguistics, polysemy sense shifts have been treated essentially as lexical rules. Jackendoff (1997, 2002) for example, integrates inferential rules into his Conceptual Structure component of the parallel language architecture. He analyzes the 'ham sandwich' metonymy as a case of 'enriched composition', where 'the ham sandwich' stands for

'the person *contextually associated with* a ham sandwich'. Other cases of enriched composition, hence comparable to metonymy in processing, are complement coercion (e.g., 'The author began the book in 1997' meaning 'the author began *writing* the book') and aspectual coercion (e.g., 'The lighthouse flashed until dawn', where a punctual predicate is coerced by the time interval adverbial to mean a series of punctual events). The common trait of these enriched composition cases is the presence of silent meaning that is not read off of syntactic structure but is ostensibly supplied by the context. Within generative approaches, enriched composition may also be processed in the syntax, as proposed by Hale and Keyser (2002) and Borer (2005). What this means is that silent syntactic structure (e.g., silent heads) carry the additional meanings, but the complete meaning is still computed compositionally.

We now turn to some experimental evidence on the processing of metonymy. The first attempts at comparing metonymic expressions with baseline literal expressions using the eye-tracking technique (Frisson & Pickering, 1999) did not find an effect of metonymic complexity in the PLACE FOR EVENT and PLACE FOR INSTITUTION sense shifts. The authors argued that the processor can obtain metonymic interpretations without appreciable difficulty. Humphrey, Kemper and Radel (2003) replicated Frisson and Pickering's (1999) processing pattern with younger as well as older adult English speakers. Recently, Rundblad and Annaz (2010) studied the development of metonymy comprehension (as compared with metaphor comprehension) in English speakers aged 5;3 to 37;1, using an off-line interpretation task. They discovered that metonymy comprehension develops over time and reaches ceiling around the age of 12 (p. 557), being significantly correlated with age and with vocabulary development. Finally, Frisson and Pickering (2007) showed that familiar metonyms were straightforward to process as tested by eye movements, but unfamiliar metonyms caused processing difficulty unless context made it

clear that the metonymic interpretation would be appropriate. We capitalize on this last finding in our research design.

Another group of experimental findings is relevant to our design as well. Pustejovsky (1995) and Jackendoff (1997) treat novel, productive metonymy on a par with other cases of enriched composition such as the complement coercion type (e.g., 'The author began the book'). These sentences cause difficulty during reading because they require costly compositional operations to repair the type mismatch between the event-selecting verb and the entity-denoting object (McElree, Traxler, Pickering, Seely & Jackendoff, 2001; Pickering, McElree & Traxler, 2005; Traxler, Pickering & McElree, 2002). For example, using a response signal speed–accuracy trade-off (SAT) task, McElree, Pylkkänen, Pickering and Traxler (2006) showed that complement coercion sentences require the deployment of additional compositional operations, in comparison with simple baseline sentences. Thus our brief overview of experimental findings suggests a possible dissociation in processing between regular, familiar metonymy associations versus more ad-hoc, novel patterns: The former do not appear to impose higher processing costs while the latter do.

Finally, let us look at what we know about cross-linguistic differences in the area of metonymy. The great bulk of the experimental and corpus work has been done on English, with the often implicit and sometimes explicit assumption that other languages work the same way and that metonymic patterns are universal. Few studies to date compare specific metonymic associations across languages. Among the rare studies that do, Brdar-Szabó and Brdar (2003) use comparable corpora in English, German, Serbo-Croatian, and Hungarian to investigate the CAPITAL FOR GOVERNMENT pattern (e.g., 'If Beijing doesn't get anything substantial from Bush at the summit in Shanghai …'). They show that this pattern is significantly more common in

English and German than in Serbo-Croatian and Hungarian. While a place name is used referentially in English and German, this metonymy is less frequently expressed with NPs in Serbo-Croatian and Hungarian. These languages substitute a locative PP and adjectival expressions instead of nominal ones. Littlemore (2009: 119), citing Woo (2008), claims that very familiar metonymies such as PRODUCER FOR PRODUCT do not work the same way in Mandarin Chinese. We have to conclude that cross-linguistic work on metonymy is very important for the various theoretical analyses described in this section, but there are not enough experimental studies to provide this information. The present study is a step down this road. To reiterate, we were interested in establishing whether the three languages of investigation, Spanish, Korean, and English, treat regular and familiar metonymic patterns such as PRODUCER FOR PRODUCT and more novel, ad-hoc metonymy such as INSTRUMENT FOR AGENT in the same way, and also whether the levels of acceptability of the two types of metonymy differ across languages.

3. Predictions

If the radical pragmatic approach is correct in the treatment of metonymy, all speaker groups should demonstrate the same accuracy and high ratings for all the metonymic patterns. This approach places concepts such as noteworthiness, centrality, and salience at the basis of metonymy: Metonymy arises because the literal and the shifted senses are centrally and saliently related to one another. There should be no language-based differences within the same metonymic pattern if the content of the clauses is the same in the three languages, as is the case in our experiment. At the same time, Nunberg's (1995) proposal that there exist language-specific polysemy conventions predicts that the same metonymic pattern may be more conventionalized within one language than within another, thus better/easier comprehended in the former language.

Cognitive linguistic approaches do acknowledge that there may be interlinguistic differences in the patterns of metonymy, but they do not provide a concrete explanation of why this should be the case. Since the inferential principles or conceptual metonymies underlying the cognitive approach are based on thought patterns and should affect all human ratiocination, it is hard to see what the basis of differences between languages could be. Therefore, we will accept that this approach to metonymy has not yet made any testable predictions of its own with respect to language variation.[2]

Finally, accounts of the generative lexicon type (Pustejovsky, Jackendoff) would easily allow for more language variation, since they consider metonymy to be based on lexical rules, and these certainly can and do vary across the languages of the world.

4. The present study

4.1 Participants

Thirty-six English, 19 Korean and 23 Spanish adult native speakers took two pen-and paper tests in their native language. The two tests were compiled in English and then translated from English into Spanish and Koreans by native speakers of the latter languages. Some of the participants were compensated with extra credit or monetarily for their time, others volunteered to take the tests.

4.2 Tests and experimental materials

In an effort to stay away from culturally-specific knowledge, we investigated only metonymic

[2] It is clear that cognitive approaches to metonymy assume Nunberg's (1995) proposal about metonymy conventions differing across languages of the world, but at least as far as we can see, there is nothing specific in their account of why these differences might come about, apart from the broad claim that metonymic inferential patterns are experience-based.

mappings that are part of global culture nowadays. We experimented with research designs and test items, piloting widely among English native speakers before translating the tests into Spanish and Korean. In order to evaluate whether the speakers access metonymic interpretations easily or not, we constructed two tasks differing in difficulty. The Paraphrase Task (see example below) asked participants to choose from four given interpretations of the sentence containing the metonymy. One paraphrase corresponded to the metonymic interpretation, one corresponded to the literal interpretation; they also had the possibility to choose both interpretations or neither interpretation. In example (1), the first choice was expected. In about 5% of the test items, (c) was the correct answer, as both the literal and the metonymic paraphrase were acceptable.

(1) Example test item from the Paraphrase Task

The first violin has the flu.

a. The first violinist is sick with the flu.

b. The first violin is not working properly because of an illness.

c. Both meanings are possible.

d. Neither meaning is possible.

In this task, we tested four metonymic patterns. Two of those we considered rarer patterns based on our own piloting: INSTRUMENT FOR AGENT (n=4) as in (1) above and another, which we termed LOOSE ASSOCIATION (n=9), comprising saliently but temporarily related things and people (e.g., 'The hamburger is a lousy tipper'). The two other patterns were considered regular and more conventionalized: POSSESSOR FOR POSSESSED (n=4) (e.g., 'Did you notice that you have a flat tire?') and PRODUCER FOR PRODUCT (n=4) (e.g., 'Proust is on the top shelf.') The test also included the same number of baseline sentences without metonymy (n=21) and 14 fillers.

The second task was an Acceptability Judgment Task, where the participants had to rate a sentence in the context of another sentence (see example in (2)). Participants had to use a scale of 1 to 5 in their rating, where 1 was 'unacceptable' and 5 was 'perfectly acceptable'.

(2) Example test items from the Acceptability Judgment Task: Acceptable metonymy

The soldiers began to move up the field during battle. The colonel gave the cannon the signal to fire.

1 2 3 4 5

In this test, we included four types of test items: baseline, experimental (good), experimental (bad) items, and fillers. The experimental sentences that were considered acceptable were divided into three metonymic patterns: INSTRUMENT FOR AGENT, LOOSE ASSOCIATION, AND PRODUCER FOR PRODUCT (n=4 in each). These were paired with sentences containing metonymies that violated metonymic requirements, see example in (3). In creating these test items, we were led by the findings of Rabagliati, Marcus and Pilkkänen (2011). Using paraphrase judgments, these researchers sought to find out whether similarity, noteworthiness,[3] cue validity (frequency of association), or salience were good predictors for the acceptability of a shifted sense. The only factor that showed some predictive effect on acceptability, and not a very robust one at that, was noteworthiness. A property is defined as noteworthy "if it offers a useful way of classifying its bearer relative to the immediate conversational interests" (Nunberg 1995: 114). The example given is that a customer is usefully classified from the point of view of a garage attendant if she says 'I am parked out back', because the attendant can identify the customer's car among others. We constructed our unacceptable test items with the intention that that they would violate this factor.

[3] The authors called this factor centrality, but they define it as noteworthiness in context where there is an important relationship between the two senses, as in Nunberg's (1995) formulation.

(3) Example test items from the Acceptability Judgment Task: Unacceptable metonymy

The small boy threw a hammer in the garage while he was playing. The workbench suffered a mental breakdown.

1 2 3 4 5

In this example, there is a meaning shift, in the sense that the workbench is a thing while suffering a mental breakdown is a characteristic of people. It is a similar linguistic process when the person who parked refers to herself but means her car. However, (3) is not a useful shift in classifying the conversational situation, it is not noteworthy. In addition to the acceptable and unacceptable metonymy items, there were 12 acceptable baseline sentences without metonymy and 14 fillers. If our research design were successful, we expected to see statistically significant differences, at a minimum, between baseline and acceptable metonymies, on the one hand, and unacceptable metonymies, on the other. We performed reliability statistics on the two tasks and obtained a Cronbach Alpha value of .791 for the Paraphrase Task and .867 on the Acceptability Judgment Task, which are highly significant statistics.

4.3 Results

4.3.1 Paraphrase Task

Recall that in the Paraphrase Task participants had to choose between four options, so chance performance is 25%. In the charts that follow, we display accuracy results (out of 1). They were obtained as follows: when the speaker chose the correct paraphrase, one point was given, in all other cases, no point was given. So that the reader can appreciate the relationship in accuracy between the two regular metonymy patterns (POSSESSOR FOR POSSESSED, PRODUCER FOR PRODUCT) and the baseline (test sentences without metonymic interpretation), we have plotted

the medians and ranges as clustered boxplots in Figure 1. The size of the boxes indicates the range of the middle 50% of values of the chosen variable, that is, the range from the 25th to the 75th quartile, or, the interquartile range. It is quite clear from the chart that there is non-normality in the data, because some outliers are shown in the English and Korean groups. Also, we see some skewness because the boxes are not symmetrical around their medians.

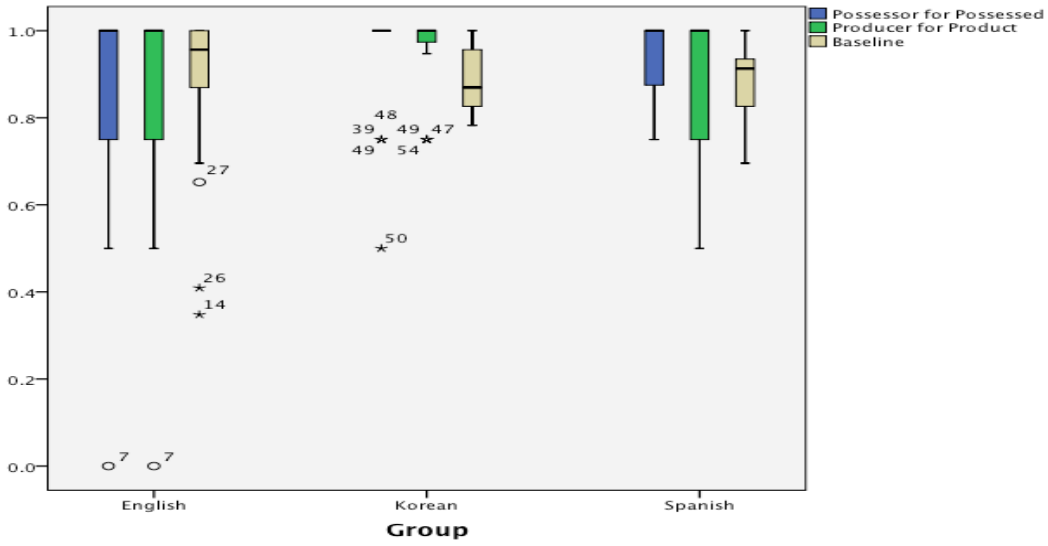

Figure 1: Clustered boxplots for the two regular metonymy conditions and the baseline in the Paraphrase Task

In Figure 2, we have plotted the same baseline accuracy contrasted with the two conditions that we stipulated to be less conventionalized, more ad-hoc: INSTRUMENT FOR AGENT and LOOSE ASSOCIATION. As can be appreciated visually, the ranges of the latter conditions in all three groups are much larger than the baseline. Furthermore, the Spanish group displays a marked accuracy difference between the ad-hoc metonymies and the baseline.

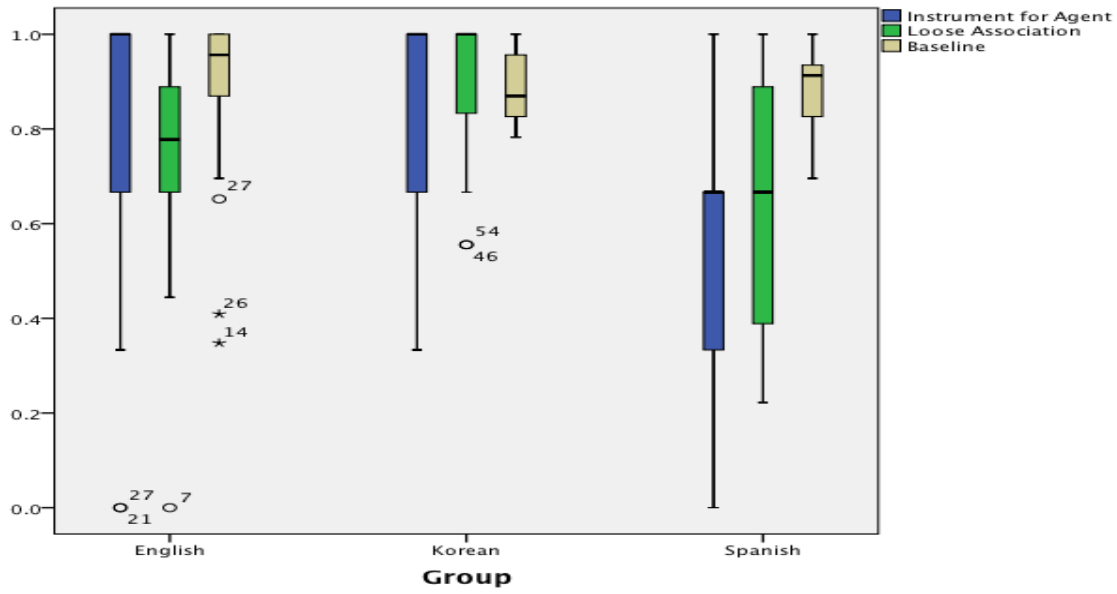

Figure 2: Clustered boxplots for the two novel metonymy conditions and the baseline in the Paraphrase Task

These visual impressions are confirmed by a one-way ANOVA on the accuracy of all conditions in the Paraphrase Task. There were significant main effects in the INSTRUMENT FOR AGENT condition ($F_{2,75} = 8.323$, $p = .001$) and in the LOOSE ASSOCIATION condition ($F_{2,75} = 6.898$, $p = .002$), but none of the other three conditions (POSSESSOR FOR POSSESSED, PRODUCER FOR PRODUCT, and Baseline) showed a condition effect ($p = .404$, $p = .313$, and $p = .848$, respectively). Levene's tests for homogeneity of variances were all non-significant, although the p values were low, so we proceeded to look at post-hoc comparisons between groups. We chose the LSD post-hoc test because it has the most power to uncover differences among groups, being the most conservative test (Larson-Hall, 2009). We present the multiple comparisons in Table 1 below.

Table 1: Multiple comparisons between means on Parapharase Task, p-values (LSD test)

Condition	*English – Korean*	*English - Spanish*	*Korean - Spanish*
INSTRUMENT FOR AGENT	.497 n.s.	.001* English > Spanish	.001* Korean > Spanish
LOOSE ASSOCIATION	.037* Korean > English	.044* English > Spanish	.0001* Korean > Spanish
POSSESSOR FOR POSSESSED	.285 n.s.	.251 n.s.	.991 n.s.
PRODUCER FOR PRODUCT	.171 n.s.	.909 n.s.	.177 n.s.
Baseline	.623 n.s.	.653 n.s.	.950 n.s.

* significant at p = .05, ">" stands for higher mean acceptance and less individual variation

We repeated the ANOVA (RM), separating what we considered the regular (POSSESSOR, PRODUCER) from the novel metonymy (INSTRUMENT, LOOSE ASSOCIATION) conditions, and compared them with the baseline. When the two regular metonymy conditions were compared to the baseline, there was no effect of condition ($F_{2,74}$ = .705, p = .497), no effect of group ($F_{2,75}$ = .577, p = .564), and no condition by group interaction ($F_{4,150}$ = 1.219, p = .305). When the two novel conditions were compared with the baseline, there was an effect of condition ($F_{2,74}$ = 19.335, p < .0001), an effect of group ($F_{2,75}$ = 8.036, p < .0001), and a condition by group interaction ($F_{4,150}$ = 5.485, p < .0001). Thus we can safely assume that the three language groups treat ad-hoc, novel metonymy differently than regular metonymy, as compared with baseline sentences without metonymy in the Paraphrase Task.

We probed further to see which individual groups treated regular and novel metonymy differently with respect to the baseline sentences. It turned out that the Spanish and the English

groups behaved similarly, in the sense that they were significantly less accurate on novel metonymy as compared with the baseline ($F_{2,21} = 18.841$, $p < .0001$ for the Spanish and $F_{2,34} = 5.499$, $p < .0001$ for the English group). The Korean group, however, judged novel and regular metonymy equally accurately to the baseline ($F_{2,17} = 1.864$, $p = .17$ for regular and $F_{2,17} = 1.272$, $p = .293$ for novel metonymy). In other words, the Koreans were more accurate on novel metonymy in comparison with the English and the Spanish groups, but all groups judged regular metonymy as accurately as they judged baseline test sentences.

4.3.2 Acceptability Judgment Task

Recall that the Acceptability Judgment Task contained context and a test sentence that had to be rated for acceptability in that context. In this task, we included experimental sentences (e.g., 'The exit row is prepared to help in case of emergency'), baseline sentences without metonymy (e.g., 'The older passengers are prepared to help in case of emergency'), as well as test sentences that violate the noteworthiness restriction on metonymy (e.g., 'The clothes dryer was not very happy'). It was hypothesized that the baseline sentences would be rated high and the unacceptable test sentences low, on a scale of 1 to 5. In the next three figures, we present clustered boxplots of each acceptable metonymy condition compared each time with baseline and unacceptable metonymy test sentences.

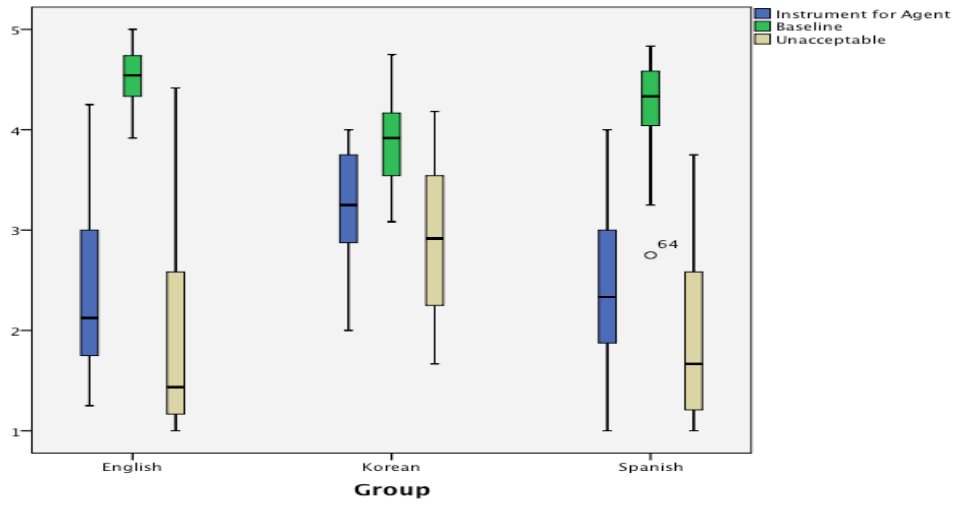

Figure 3: Clustered boxplots of the INSTRUMENT FOR AGENT metonymy compared with baseline and unacceptable metonymy in the judgment task

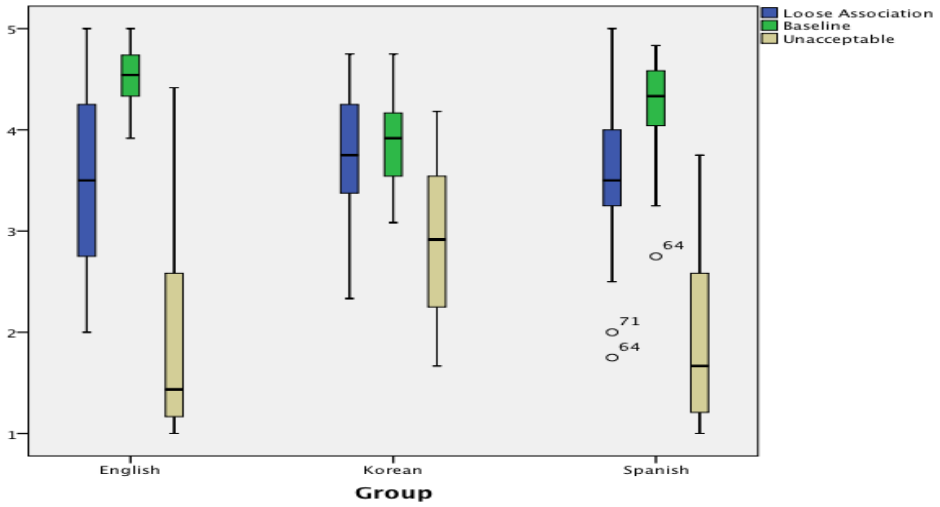

Figure 4: Clustered boxplots of the LOOSE ASSOCIATION metonymy compared with baseline and unacceptable metonymy in the judgment task

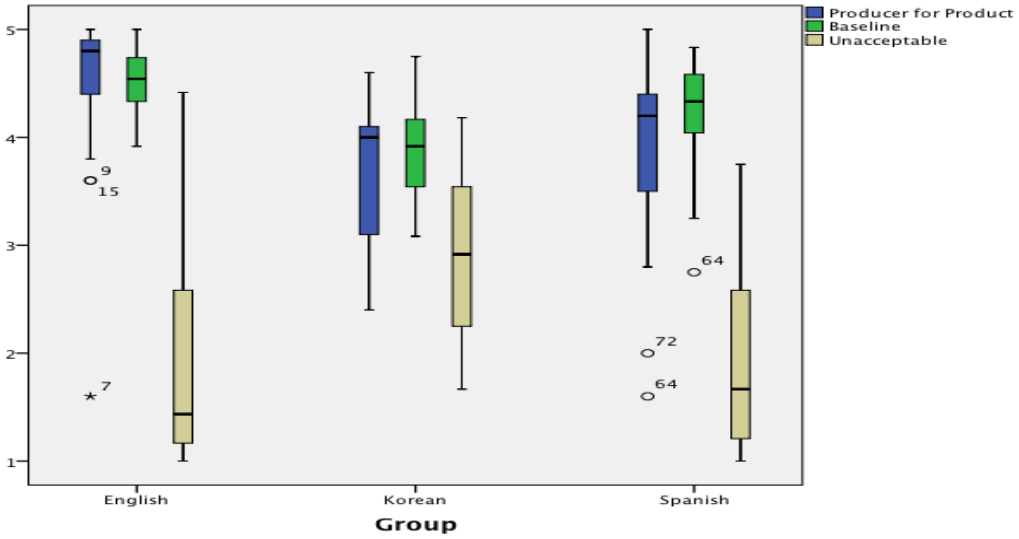

Figure 5: Clustered boxplots of the PRODUCER FOR PRODUCT metonymy compared with baseline and unacceptable metonymy in the judgment task

An ANOVA with repeated measures on the three metonymy conditions in this task reveals a significant effect of condition ($F_{2, 74} = 96.289$, $p < .0001$), no effect of group ($F_{2, 75} = 1.961$, $p = .148$), and a significant condition by group interaction ($F_{4, 150} = 12.698$, $p < .0001$). These results suggest that the three groups do not treat all the metonymy conditions equally. A one-way ANOVA on all the conditions shows a main effect in all conditions but LOOSE ASSOCIATION. There were significant main effects in the INSTRUMENT FOR AGENT condition ($F_{2,75} = 8.325$, $p = .001$), the PRODUCER FOR PRODUCT condition ($F_{2,75} = 10.333$, $p = .0001$), and in the LOOSE ASSOCIATION condition ($F_{2,75} = 6.898$, $p = .002$), the baseline condition ($F_{2,75} = 14.618$, $p = .0001$), and the unacceptable metonymy condition ($F_{2,75} = 7.824$, $p = .001$). Since the assumption for equal variances was confirmed by the Levene statistic, we further looked at LSD post-hoc comparisons. The differences between groups are presented in Table 2.

Table 2: Multiple comparisons between means on the judgment task, p-values (LSD test)

Condition	*English – Korean*	*English - Spanish*	*Korean - Spanish*
INSTRUMENT FOR AGENT	.001* Korean > English	.598 n.s.	.0001* Korean > Spanish
LOOSE ASSOCIATION	.319 n.s.	.785 n.s.	.253 n.s.
PRODUCER FOR PRODUCT	.0001* English > Korean	.001* English > Spanish	. 396 n.s.
Baseline	.0001* English > Korean	.016* English > Spanish	.006* Spanish > Korean
Unacceptable metonymy	.0001* Korean > English	.853 n.s.	.002* Korean > Spanish

* significant at $p < .05$, ">" stands for higher mean acceptance and less individual variation

Perhaps even more interesting from the point of view of our research questions, we would like to know not only how the groups differ from each other but also what the variation is within groups. In order to uncover whether each individual metonymy differs from the baseline and the unacceptable test sentences within each language, we performed a series of paired samples t-tests. The results are shown in Table 3. Ideally, we would like all experimental metonymy conditions to be rated the same as the baseline conditions but differently from the unacceptable ones. However, that is not exactly what we find. While - as Table 3 reveals - all groups rate the experimental metonymies significantly higher than the unacceptable metonymies,

validating our test instrument, comparisons between the experimental metonymies and the baseline bring to light some unexpected outcomes. Furthermore, the English group rates only the PRODUCER FOR PRODUCT metonymy as highly as the baseline test sentences. The Korean group rates both PRODUCER FOR PRODUCT and LOOSE ASSOCIATION as highly as the baseline. Finally, the Spanish speakers situate the ratings of all metonymy conditions somewhere between the acceptable baseline and the unacceptable metonymy, with significant differences everywhere.

Table 3: Differences between metonymy experimental conditions, baseline, and unacceptable metonymy for all language groups

Group	Contrast	P value
English NSs		
	INSTRUMENT FOR AGENT − Unacceptable	.0001
	INSTRUMENT FOR AGENT − Baseline	.0001
	PRODUCER FOR PRODUCT − Unacceptable	.0001
	PRODUCER FOR PRODUCT − Baseline	.8 n.s.
	LOOSE ASSOCIATION − Unacceptable	.0001
	LOOSE ASSOCIATION − Baseline	.0001
Korean NSs		
	INSTRUMENT FOR AGENT − Unacceptable	.046
	INSTRUMENT FOR AGENT − Baseline	.001
	PRODUCER FOR PRODUCT − Unacceptable	.0001
	PRODUCER FOR PRODUCT − Baseline	.172 n.s.
	LOOSE ASSOCIATION − Unacceptable	.0001
	LOOSE ASSOCIATION – Baseline	.502 n.s.
Spanish NSs		
	INSTRUMENT FOR AGENT − Unacceptable	.026
	INSTRUMENT FOR AGENT − Baseline	.0001
	PRODUCER FOR PRODUCT − Unacceptable	.0001
	PRODUCER FOR PRODUCT − Baseline	.023
	LOOSE ASSOCIATION − Unacceptable	.0001
	LOOSE ASSOCIATION − Baseline	.0001

Note: All significant differences are in the expected directions. Metonymy test sentences are evaluated as more acceptable than unacceptable test sentences but less acceptable than the baseline sentences.

5. Discussion

Let us summarize the results of the two experimental tests on the comprehension and acceptability of novel and regular metonymic patterns by native speakers of three languages: English, Korean, and Spanish. The first impression on looking at the boxplots in Figures 1 through 5 is that there is a lot of variation in the metonymy data of the native speakers, a lot

more variation than is common when native speakers are experimentally tested on syntactic or semantic properties. Standard deviations on both tests are quite dramatic. This fact on its own suggests that there is instability in judging and comprehending such test sentences, and therefore large inter-speaker differences. While some individuals are quite accepting of metonymic patterns, other individuals judge them more conservatively as relatively unacceptable. We have to emphasize that the theoretical literature on metonymy is quite oblivious to this individual variation, even within the metonymic patterns that are considered the most highly conventionalized (see e.g., Figure 1). It is clear that shifting a meaning, even if it is a regular linguistic process, is not without its processing costs, as attested by individual variation and lower accuracy. In order to substantiate this finding, further processing investigations of all metonymic patterns are warranted.

A salient and significant finding of this experiment is the different treatment of regular and novel metonymic patterns by the speakers of all three languages tested. This finding may be universal, especially when we consider the fact that the languages we tested are typologically distinct. We can expect to see regular-novel metonymy divergence repeated in language after language. This is suggested by the main effect of condition on the RM ANOVAs for both tests. For example, in the Paraphrase Task, when the two regular metonymy conditions were compared with the baseline, there was no effect of condition; while, when the two novel conditions were compared with the baseline, there was a significant effect of condition. These results were not unexpected. The literature on metonymy processing (Frisson & Pickering, 1999, 2007) has documented this possible differentiation of familiar and non-familiar metonymic patterns in English. The contribution of the present experimental study is to confirm that such differentiation exists in Spanish and Korean, as well. At the same time, we can fully expect that the levels of

conventionalization of the various metonymic patterns will vary in language after language.

How can the linguistic approaches to metonymy account for this differentiation? It seems to us that both radical pragmatic approaches and generative lexicon approaches can easily accommodate such findings. Within radical pragmatics, the different levels of conventionalization of the metonymic patterns explain the differential accuracy in their comprehension. Nunberg (1995: 116) classifies the two types of metonymy as 'regular polysemy' or 'systematic polysemy' versus 'occurrent metonymies', the latter being such meaning shifts "where the relation between two domains can only be exploited in a restricted range of situations." For example, the 'ham sandwich' designation is useful when the discourse situation is directly concerned with customers and ordering, but it is unlikely that a waiter is going to refer to a customer in this way when referring to some other action outside of the ordering situation (e.g., 'I saw the ham sandwich at table 7 the other day driving a Mercedes', Nunberg's example). While both types of meaning transfer are subject to the general condition of noteworthiness, the occurrent metonymies (what we have dubbed novel or ad-hoc metonymies in this article) are not likely to be lexicalized processes, according to Nunberg. Our findings are certainly supportive of this view.

The generative lexicon approach (Copestake & Briscoe, 1995; Jackendoff, 1997; Pustejovsky, 1995) proposes the treatment of metonymies largely as lexical rules. The silent meaning of the shifted sense is introduced in the sentence computation by a purely semantic rule called type-shifting. More specifically, the interpretation of 'The writer began the book' is coerced into an event reading (writing the book) to satisfy the predicate's semantic requirements. However, this coercion process is dependent, in the case of complement coercion, on the idiosyncratic properties of verbs like 'begin', 'start', etc. The cases of regular metonymy are

treated as lexical processes, or rules, with semantic import over the meaning of the whole sentence. With respect to novel, ad-hoc metonymy (the 'ham sandwich' example derivation), Jackendoff (1997: 58) postulates that a speaker's knowledge includes a principle of enriched composition: "Interpret an NP as [PERSON CONTEXTUALLY ASSOCIATED WITH NP]." Thus, this view can certainly accommodate the regular versus novel metonymy distinction present in our data.

The next research question this experimental study asked is whether there are differences among languages (English, Korean, and Spanish) with respect to the acceptability of novel versus regular metonymy. The existing literature on cross-lingual differences in metonymy acceptability (e.g., Kamei & Wakao, 1992; Peters, 2003) looks for metonymies in corpora and is concerned with their machine translation into other languages, which is why it mostly treats regular metonymy. To the best of our knowledge, the present study is the first to compare novel, productive metonymy across languages. Comparisons between acceptability in the three languages, as reflected in Table 1 for the Paraphrase Task, reveal that speakers of our three languages do not differ that much in the comprehension of regular metonymy, but they do treat novel metonymy differently. Language comparisons in the Acceptability Judgment Task are more difficult to evaluate (see Table 2), but generally speaking, differences are uncovered in the INSTRUMENT FOR AGENT condition and the PRODUCER FOR PRODUCT condition. The former differences are due to the Korean speakers accepting this type of metonymy more readily than the Spanish and English speakers. The latter differences are due to the English speakers accepting PRODUCER FOR PRODUCT sentences much more readily than Korean and Spanish speakers. If we accept that the Acceptability Judgment Task was the more reliable one in terms of speakers' underlying language competence, since sentences appeared in context and ratings

were based on a scale, then the boxplots in Figures 3-5 reveal significant differences in the way the three languages treat metonymy. Koreans are more liberal in the treatment of loosely contextually associated, novel metonymic shifts; however, they treat regular metonymy more or less in the same way (see Table 3), most likely computing the meaning shifts online without much conventionalization, as reflected in large standard deviations and variability among individual speakers. English speakers make the clearest distinction between regular and novel metonymy, with regular metonymy lexicalized and novel metonymy computed online. Finally, the Spanish speakers are the most conservative from the point of view of lexicalizing the metonymic patterns: They treat regular metonymy much more like novel metonymy, in comparison with the English speakers. Perhaps the most surprising finding is the Korean speakers' high acceptance of the LOOSE ASSOCIATION pattern, indicative of their more productive treatment of novel metonymy than the English speakers.

How can the accounts discussed above accommodate the intra-lingual findings? Within generative lexicon approach, conventionalized and language-specific aspects of the general processes of meaning transfer are expressed as lexical rules. Since these rules can be postulated to serve as language-specific filters on the general, universally available cognitive process, this approach can easily account for the findings. The radical pragmatics approach does postulate that differences between languages exist, but they are considered to be experience-based. For example, Nunberg (1995: 118), in discussing examples of language-specific rules, mentions that French speakers use the names of fruits to refer to the brandies made from them (*une prune* 'a plum', *un poire* 'a pear') while this meaning shift is unavailable in English. The implication is that the cultural experience of French people with various brandies allows for the meaning shift. However, in our test sentences, we only included situations that are not culture-sensitive and

specifically avoided situations that could pertain to one culture but not the others. This is not difficult to do in our globalized world. Since we still obtained language disparities, we attribute them to the working of the various rules that serve as language-specific filters (Copestake & Briscoe, 1995).

Finally, within cognitive approaches to metonymy, many linguists are explicitly mindful of cross-linguistic differences in the expression of, for example, paragon names (Barcelona, 2004), predicative adjectives (Brdar-Szabo & Brdar 2004), or place names (Brdar & Brdar-Szabo 2009), among others. More generally speaking, chapters in Panther, Thornburg and Barcelona (2009) argue for figurative thought, that is, metaphor and metonymy, partially motivating grammatical structure. On the other hand, Brdar-Szabo and Brdar (2004: 327) submit that whether a language employs certain metonymic processes may depend on the shape of its current grammar. Unfortunately, we cannot answer this chicken-or-egg issue here with our experimental design. We would not speculate as to the reasons why, for example, Korean seems to be more accepting of novel metonymy than English is. A plausible answer to such a question can only come from a much wider comparative study of the grammar, cognitive processes, and metonymy patterns among languages. However, we would like to argue that a comprehensive theory of metonymy should be able to explain and even predict experimental findings of cross-linguistic differences. Such a theory should take into account both regular and novel, productive metonymic patterns, their levels of conventionalization, as well as the way they are processed.

It is important to remind the reader that our discussion of processing costs related to novel and regular metonymy computation is only tentative, since our measures in this experimental study were two off-line interpretation tasks. We have used accuracy and individual variation as suggestive of on-line computation and processing costs. However, in order to answer

the processing resources question directly, researchers need to look next at the online comprehension of such sentences in various languages, using measures such as self-paced reading, eye-tracking, and event-related brain potentials.

In conclusion, in this article we presented results of two off-line comprehension tasks, investigating the acceptability of novel and regular metonymy by speakers of English, Korean, and Spanish. We were interested in uncovering regular-novel metonymy computation discrepancies, and whether they are treated differently in the different languages. The distinction between novel and regular metonymy is not new and is discussed at length by the existing theoretical treatments of metonymy. We easily obtained further experimental support for this distinction. Furthermore, and more importantly, we discovered that the speakers of the three languages treat novel and regular metonymy differently. Significant findings were the acceptability of novel metonymy in Korean and the lack of conventionalization effect for regular metonymy in Korean and Spanish. We would like to highlight the importance of cross-lingual examination of metonymic patterns, not only in comprehension but in processing as well.

Acknowledgements:

The authors would like to acknowledge the participants in this experimental study, as well as Prof. Park Yeon-Mi at Han-Kyung University, Prof. Kim Young-Hwa at Hallym University, Alice Davison at the University of Iowa, and Jason Rothman at the University of Florida (at the time) for help with recruiting participants. Gonzalo Campos and Diego Pascual y Cabo helped with the translation of the Spanish tests, thank you! The editor provided valuable comments and insight, for which we are grateful.

References:

Barcelona, A. (2011). Reviewing the properties and prototype structure of metonymy. In: Benczes, R., Barcelona, A., Ruiz de Mendoza Ibáñez, F. J. (eds.) *Defining Metonymy in Cognitive Linguistics: Towards a Consensus View* (pp. 7-58). Amsterdam: John Benjamins, Amsterdam.

Barcelona, A. (2004). Metonymy behind grammar: the motivation of the seemingly "irregular" grammatical behavior of English paragon names. In Radden, G., Panther, K.-U. (eds.), *Studies in Linguistic Motivation* (pp. 357-374). Berlin: Walter de Gruyter.

Benczes, R., Barcelona, A., & Ruiz de Mendoza Ibáñez, F. J. (eds.) (2011). *Defining Metonymy in Cognitive Linguistics: Towards a Consensus View*. Amsterdam: John Benjamins.

Borer, H. (2005). *Structuring Syntax. An Exo-skeletal Trilogy* (vols. 1-2). Oxford: Oxford University Press.

Brdar-Szabó, R., & Brdar, M. (2003). Referential metonymy across languages: What can cognitive linguistics and contrastive linguistics learn from each other? *International Journal of English Studies 3*, 85-105.

Brdar-Szabó, R., & Brdar, M. (2004). Predicative adjectives and grammatical-relational polysemy: The role of metonymic processes in motivating cross-linguistic differences. In Radden, G., Panther, K.-U. (eds.), *Studies in Linguistic Motivation* (pp. 321-356). Berlin: Walter de Gruyter.

Brdar, M., & Brdar-Szabó, R. (2003). Metonymic coding of linguistic action in English, Croatian and Hungarian. In: Panther, K.-U., Thornburg, L. (eds.), *Metonymy and Pragmatic Inferencing* (pp. 241-266). Amsterdam: John Benjamins.

Brdar, M., & Brdar-Szabó, R. (2009). The (non-)metonymic use of place names in English, German, Hugrarian and Croatian. In Panther, K.-U., & Thornburg, L., & Barcelona, A. (eds.), *Metonymy and Metaphor in Grammar,* (pp. 229-258). Amsterdam: John Benjamins.

Brennan, J., & Pylkkänen, L. (2008). Processing events: behavioral and neuromagnetic correlates of aspectual coercion. *Brain and Language, 106*, 132-143.

Copestake, A., Briscoe, E. J. (1995). Semi-productive polysemy and sense extension. *Journal of Semantics, 12*, 15-67.

Croft, W. (2002). The role of domains in the interpretation of metaphors and metonymies. In Dirven, R. & Pörings, R. (eds.), *Metaphor and Metonymy in Comparison and* Contrast (pp. 161–205). Berlin/New York: Mouton de Gruyter,.

Dölling, J., (1995). Ontological domains, semantic sorts and systematic ambiguity. *International Journal of Human-Computer Studies, 43*, 785–807.

Fauconnier, G. (1985). *Mental Spaces*. Cambridge, MA: MIT Press.

Frisson, S., & Pickering, M. (1999). The Processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25* (6), 1366-1383.

Frisson, S., & Pickering, M. (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* (4), 595-613.

Gibbs, R. (2009). A psycholinguist's view on cognitive linguistics: An interview with Ray W. Gibbs. *Annual Review of Cognitive Linguistics, 7*, 302-318.

Hale, K., & Keyser, J. (2002). *Prolegomena to a Theory of Argument Structure*. Cambridge, MA: MIT Press.

Harris, J., Pylkkänen, L., McElree, B., & Frisson, S. (2008). The cost of question concealment: Eye-tracking and MEG evidence. *Brain and Language 107*, 44-61.

Humphrey, H., Kemper, S., & Radel, J. (2004). The time course of metonymic language text processing by older and younger adults. *Experimental Aging Research: An International Journal Devoted to the Scientific Study of the Aging Process, 30* (1), 75-94

Jackendoff, R. (1997). *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford, UK: Oxford University Press.

Kamei, S.-I., & Wakao, T. (1992). Metonymy: Reassessment, Survey of Acceptability, and its Treatment in a Machine Translation System. In: H. Thompson (ed.), *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics* (pp. 309-311). Stroudsburg, PA: The Association for Computational Linguistics.

Kövecses, Z., & Radden, G. (1998). Metonymy: Developing a cognitive linguistic view. *Cognitive Linguistics, 9*, 37–77.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind,* Chicago: The University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: The University of Chicago Press.

Larson-Hall, J. (2009). *A Guide to Doing Statistics in Second Language Research Using SPSS*. New York/London: Routledge.

Lascarides, A., & Copestake, A. (1998). Pragmatics and word meaning. *Journal of Linguistics,*

34, 387-414.

Littlemore, J. (2009). *Applying Cognitive Linguistics to Second Language Learning and Teaching.* Basingstoke: Palgrave Macmillan.

Markert, K., & Nissem, M., 2003. Corpus-based metonymy analysis. *Metaphor and Symbol, 18* (3), 175-188.

McElree, B., Pylkkänen, L., Pickering, M., & Traxler, M. (2006). The time course analysis of enriched composition. *Psychonomic Bulletin & Review, 13* (1), 53-59.

McElree, B., Traxler, M., Pickering, M., Seely, R., Jackendoff, R. (2001). Reading time evidence for enriched composition. *Cognition, 78*, B17-B25.

Murphy, G. (1997). Polysemy and the creation of novel word meanings. In: Ward, T.B., Smith, S., & Vaid, J. (eds.), *Creative Thought: An Investigation of Conceptual Structures and Processes* (pp. 235-265). Washington DC: American Psychological Association.

Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy, 3*, 143-184.

Nunberg, G. (1995). Transfers of meaning. *Journal of Semantics, 12*, 143-184.

Nunberg, G., & Zaenen, A. (1992). Systematic polysemy in lexicology and lexicography. In: Tommola, H., Varantola, K., Salmi-Tolonen, T., & Schopp T. J. (eds.), *Proceedings of Euralex II* (pp. 387-396). Swansea, UK: Euralex.

Ostler, N., & Atkins, B.T.S. (1992). Predictable meaning shift: Some linguistic properties of lexical implication rules. In: Pustejovsky, J., & Bergler, S. (eds.), *Lexical Semantics and Knowledge Representation* (pp. 87-100). Berlin: Springer Verlag.

Panther, K.-U., & Thornburg, L. (eds.) (2003). *Metonymy and Pragmatic Inferencing*. Amsterdam: John Benjamins.

Panther, K.-U., & Thornburg, L. (2007). Metonymy. In Geeraerts, D., Dirven, R., & Cuyckens, H. (eds.), *Handbook of Cognitive Linguistics* (pp. 236-263). Oxford: Oxford University press.

Panther, K.-U., & Thornburg, L., & Barcelona, A. (eds.) (2009). *Metonymy and Metaphor in Grammar*. Amsterdam: John Benjamins.

Papafragou, A. (1996). On metonymy. *Lingua, 99*, 169-195.

Peters, W. (2003). Metonymy as a cross-lingual phenomenon. In: Barnden, J., Glasbey, S., Lee, M., & Wallington A. (eds.), *Proceedings of the ACL 2003 Workshop on Lexicon and Figurative Language* (pp. 1-9). Stroudsburg, PA: Association for Computational Linguistics.

Pickering, M., McElree, B., & Traxler, M. (2005). The difficulty of coercion: A response to de Almeida. *Brain & Language, 93*, 1-9.

Piñango, M.M., Zurif, E., & Jackendoff, R. (1999). Real-time processing implications of enriched composition and the syntax-semantics interface. *Journal of Psycholinguistic Research, 28*, 395–414.

Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics, 17* (4), 409-441.

Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.

Pylkkänen, L. (2008). Mismatching meanings in brain and behavior. *Language and Linguistics Compass 2,* (4), 712–738.

Rabagliati, H., Marcus, G., & Pylkkänen, L. (2011). Rules, radical pragmatics, and restrictions on regular polysemy. *Journal of Semantics, 28* (4), 485-512.

Rundblad, G., & Annaz, D. (2010). Development of metaphor and metonymy comprehension: receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology, 28* (3), 547-563.

Traxler, M., McElree, B., Williams, R., & Pickering, M. (2005). Context effects in coercion: Evidence from eye movements. *Journal of Memory and Language, 53*, 1-25.

Traxler, M., Pickering, M., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language, 47*, 530-547.

Woo, I. (2008). Understanding Metaphor: Taiwanese Students and English Language Metaphor. MA thesis, University of Birmingham.