

Journal of Semantics: 1–27
doi:10.1093/jos/ffi005

Rules, Radical Pragmatics and Restrictions on Regular Polysemy

HUGH RABAGLIATI

Brown University and New York University

GARY F. MARCUS

New York University

LIINA PYLKKÄNEN

New York University

Abstract

Although regular polysemy [e.g. producer for product (*John read Dickens*) or container for contents (*John drank the bottle*)] has been extensively studied, there has been little work on why certain polysemy patterns are more acceptable than others. We take an empirical approach to the question, in particular evaluating an account based on rules against a gradient account of polysemy that is based on various radical pragmatic theories (Fauconnier 1985; Nunberg 1995). Under the gradient approach, possible senses become more acceptable as they become more closely related to a word's default meaning, and the apparent regularity of polysemy is an artefact of having many similarly structured concepts. Using methods for measuring conceptual structure drawn from cognitive psychology, Study 1 demonstrates that a variety of metrics along which possible senses can be related to a default meaning, including conceptual centrality, cue validity and similarity, are surprisingly poor predictors of whether shifts to those senses are acceptable. Instead, sense acceptability was better explained by rule-based approaches to polysemy (e.g. Copestake & Briscoe 1995). Study 2 replicated this finding using novel word meanings in which the relatedness of possible senses was varied. However, while individual word senses were better predicted by polysemy rules than conceptual metrics, our data suggested that rules (like producer for product) had themselves arisen to mark senses that, aggregated over many similar words, were particularly closely related.

1 INTRODUCTION

Although most words have relatively stable form representations, their interpretations change frequently to suit the context. Particularly interesting in this regard is the phenomenon of regular polysemy, in which similar alternations between senses extend across a whole class of lexical items (see Table 1). For instance, words for producers can also be

Regular polysemy	Default sense example	Shifted sense example	Generalizations
Object–abstract object	The shiny DVD	The hour-long DVD	CD/record/film /disc etc.
Container–contents	The broken pot	The boiling pot	pan/dish/kettle /cauldron etc.
Institution–person	The large nunnery	The friendly nunnery	school/university /church/bakery etc.
Animal–food	The noisy chicken	The tasty chicken	rabbit/lamb/penguin /sasquatch etc.

Table 1 Examples of regular polysemy

used for their products, as in, *thirsty/magnificent Beethoven, hungry/valuable Picasso, drunken/epic Hemingway*.

Regular polysemy is typically modelled as the application of lexical rules, like producer–product, which take words from within one particular semantic field and shift that sense to another (Ostler & Atkins 1992; Copestake & Briscoe 1995; Pustejovsky 1995; Murphy 1997). Alternatively, such rules have also been proposed to operate within the syntax (Borer 2005) or during semantic composition (Dölling 1995). As a psychological explanation, these rules have certain advantages; in particular, they account for the generative nature of regular polysemy, avoid the need to list predictable senses in the mental lexicon and provide a good analogy with morphological rules. But while researchers have made progress in providing accounts of *how* regular polysemy operates, as well as documenting different types of regular polysemy, rule-based theories have made little progress in explaining *why* particular regularities exist. Most models of regular polysemy are purely descriptive, and do not try to account for why it is possible to use a producer to stand for its product but not to use a product to stand for its producer (e.g. *the 9th symphony was deaf* cannot describe Beethoven’s hearing loss).

Here, we take an experimental approach to explaining why certain shifts between senses are more acceptable than others. In particular, we test whether discrete rules are necessary to explain regular polysemy or whether seemingly rule-based shifts could be explained by a more gradient account, one in which senses are instead assigned based on whether they are closely related to the default meaning of a word.

More specifically, we test the predictions of a variety of radical pragmatic conceptualizations of word meaning (Fauconnier 1985; Nunberg 1995; Papafragou 1996; Strigin 1998). These theories have been among the few to provide justifications for why only certain shifts

between senses are permissible. Lexical entries are not modified by rules operating over specific links between semantic classes (e.g. between containers and contents) but instead changes in sense occur because world knowledge is used to infer which meanings are plausible for a word. For instance, Nunberg (1995) argued that a word's sense could be shifted from one meaning to another whenever the latter became particularly 'noteworthy' in the context, that is, when there is an important relation between the two. A sentence like the *hour-long DVD* is felicitous because movies are noteworthy in the context of DVD discs (DVDs are made to store and play movies, they have pictures of movies on their covers, etc.). However DVD discs are not particularly noteworthy when thinking about movies (unlike actors, directors, etc.); hence, *the shiny movie* is unlicensed. Another example can be seen in the acceptability of *Plato is on the top shelf* (Fauconnier 1985). There is an important link between Plato and his works, which allows the former to stand for the latter; however, the link from Plato's works back to the author is not as strong. Rather, the most obvious links are to various philosophical ideas, and so *The Republic died in Athens* does not have an interpretation where *The Republic* refers to Plato.

Under radical pragmatic theories, many of the sense shifts that are typically explained by regular polysemy rules (like producer-product or object-abstract object) are just epiphenomena of a set of concepts being similarly structured, resulting in similar senses. Just as DVDs are closely associated with movies, so a CD will be closely associated with the music it contains (*the classical CD*) and a book will be closely associated with its story (*the thrilling book*). Similarly, a shift from producers to products will also appear to be a rule because the people it applies over (writers, composers, artists) are all intimately associated with their work. The apparently regular and rule-governed nature of lexical semantics results from regularities in the mental representation of the external world. That said, these theories do also have a place for mechanisms akin to polysemy rules. For example, Nunberg (1995) acknowledges that conventional, language-specific polysemy patterns exist. But these conventions are implemented by the same noteworthiness-based mechanism, and so his theory predicts that world knowledge should affect sense acceptability independently of any rule.

Within radical pragmatics, disputes arise as to which aspects of world knowledge define sense acceptability. As mentioned, Nunberg (1995) argues that senses are assigned based on noteworthiness: if another meaning is noteworthy for a word given the context, that meaning is assignable. Researchers in the relevance theory tradition have made proposals along similar lines, advocating that senses are licensed when they are relevant to the current situation (Papafraou

1996; Wilson 2003).¹ Finally, Fauconnier (1985) argues that one meaning can be shifted to another if the two meanings are connected within an ‘idealized cognitive model’, a model of conceptual representation drawn from cognitive linguistics. Such connections signal important relations between concepts, which can arise for a variety of reasons (e.g. cultural importance). The ability to set up a new connection and therefore a new sense will depend on the importance of that relation.

Although notions like noteworthiness and relevance are intuitively appealing explanations for why we can talk about *the hour-long DVD* but not *the shiny movie*, what they describe is not entirely clear. One possibility is that they can be defined in terms of relatively low-level metrics. In previous work, Nunberg (1979) had suggested that the cue validity of the default sense for the shifted sense might be critical. The notion of cue validity is derived from Eleanor Rosch’s work on categorization (e.g. Rosch & Mervis 1975), and describes the conditional probability of an entity being in a category given that it has a feature, (e.g. the probability that an object is an elephant given that it has tusks). Of course, senses should not be thought of as features, for instance, a paraphrase of *the hour-long DVD* is not ‘the hour-long contains a movie’. Instead, Nunberg’s cue validity should be thought of as the probability of one concept (a movie) occurring in the context of another (a DVD), independently of why they are associated (i.e. no matter whether the DVD contains the movie, is used for pirating the movie, etc.). This matches our intuitions regarding our current polysemy examples: given a DVD disc, it is relatively predictable that it will be somehow associated with a movie, but given a movie one might not want to predict that it is associated with a DVD. We therefore tested whether cue validity predicted sense acceptability.

But noteworthiness, relevance and Fauconnier’s links seem to reflect a higher level of conceptual structure than simple cue validity. What all these theories have in common is the assumption that senses become licensed when they are somehow intrinsically related, that is to say secondary senses have to play some important role in our conceptualization of the first sense. By contrast, cue validity allows senses to be related for accidental or extrinsic reasons.

Intrinsic relatedness can be well expressed and operationalized by the notion of ‘conceptual centrality’ (Murphy & Medin 1985; Ahn 1998; Sloman *et al.* 1998). Taken from cognitive psychology, conceptual centrality describes the extent to which one aspect of a concept causes

¹ Although what makes a sense relevant differs between theories.

other features to come about. For instance, the fact that animals have DNA causes a range of other animal properties, so DNA is conceptually central for a modern-day understanding of animals. In terms of polysemy, conceptual centrality can be thought of as the extent to which the secondary sense causes important features of the primary sense, for example, we may be licensed to assign the word *DVD* a movie sense because DVDs are designed to play movies, and therefore movies necessitate various facts about DVDs and their design.

The importance of central features has been emphasized in various theoretical accounts of conceptual representation (Murphy & Medin 1985; Sloman *et al.* 1998; Ahn *et al.* 2000; Rehder 2003), and confirmed by a number of experimental tests (Sloman *et al.* 1998; Ahn *et al.* 2000; Rehder & Hastie 2004; Rehder & Kim 2006). Central features play a particularly prominent role in knowledge-based accounts of concepts (Murphy & Medin 1985). In this account, the coherence of a concept (similar to its naturalness) is determined by the extent to which it can be explained via an intuitive theory. Central features by definition cause (and thus explain) a large number of other features, and so provide the glue that makes a concept cohere. If word meanings reflect our concepts, then central features that are important for characterizing a concept should also be important for characterizing the senses associated with it.² We therefore tested whether conceptual centrality accurately predicted sense acceptability.

Of course, it could be that sense acceptability does not depend on concepts being intrinsically related, and instead, the secondary sense just has to be particularly prominent when thinking of the default sense. In her relevance-theoretic account of sense change, Papafragou (1996) argues along these lines and proposes that relevant senses are determined by salience. Salience describes whether a feature is ‘attention grabbing’, independent of whether it is frequent or central. For example, the fact that sharks eat people is salient, even though it is rare and necessitates little else about sharks. Under this account, when thinking of the word *DVD*, movie would be a salient concept, which would make a movie sense licensed. We therefore examined whether salience predicts sense acceptability.

The theories discussed above have assumed that word senses are related, but not necessarily overlapping, aspects of conceptual structure. For example, while DVDs are related to movies, they do not share

² Indeed, Sloman *et al.* (1998) have argued that central features are critical for metaphors, inasmuch as they resist transformation and so constrain which interpretations are possible, which provides some indirect support for this idea.

many features. By contrast, an alternative assumption in polysemy is that word senses *do* overlap in meaning, that is to say they are similar. Similar concepts share structure and features (e.g. a volcano and a mountain are both similar and related, while a volcano and a volcanologist are related but not similar), and similarity is critical for theories of conceptual representation. For instance, people categorize similar items together and dissimilar items apart and are more likely to make property inductions from one item to a similar alternative.

Despite its importance for the psychology of concepts, similarity seems an intuitively implausible predictor of sense acceptability (movies and DVDs are clearly not similar yet *hour-long DVD* is licensed). Nevertheless, certain senses do appear to be similar (such as the different senses of the words *ring* and *line*), and various theories have argued that similarity licenses senses. For instance, Apresjan (1974) explicitly proposes that polysemous senses are similar to one another, while Gibbs *et al.* (1994) argue that different senses of the word *stand* reflect meanings that are similar to one of a small number of prototypical meanings.

Experimental evidence, however, tends not to support these theories. Klein & Murphy (2001, 2002) demonstrated that the typical effects of similarity do not hold between polysemous senses: different senses are not rated as particularly similar, are not categorized together and do not license inductive inferences from one to the other (e.g. if wrapping paper increases in cost, subjects do not infer that a liberal paper will increase in cost). All of this is inconsistent with representational overlap between senses; indeed, Klein & Murphy (2002) endorse the idea that senses are related but not similar.

Still, there are reasons to pause before fully accepting the conclusion that similarity plays little role in polysemy. First, Klein and Murphy studied relations between acceptable senses, rather than contrasting acceptable and unacceptable senses. Similarity may still be important for determining whether a sense is licensed, even if it has little effect on processing within licensed senses. Second, their materials mixed different types of polysemy. Some senses were regular (e.g. *noisy/tasty chicken*) but others were irregular [e.g. the different senses of *trunk* (of a car/similar to a suitcase) do not represent one instance of a lexical generalization that cuts across English]. It may be that similarity is unimportant for irregular polysemy, but important for regular polysemy, where new senses often have to be generated.

In summary, the present study investigates whether restrictions on polysemous senses can arise from the graded structure of our

conceptual representations. We test whether the acceptability of a sense increases as it becomes more related to the default meaning, via either conceptual centrality, cue validity or salience, or alternately whether senses become licensed if they are conceptually overlapping, that is, are similar. These possibilities were contrasted with the hypothesis that sense acceptability is determined by whether or not a word's default meaning falls within the semantic field of a licensed polysemy rule.

Study 1 assessed whether relatedness or rules better explained word sense acceptability for a large set of extant English words. Using word pairs, we recorded how willing participants were to let one word be used in place of its pair (e.g. use the word *DVD* to mean movie) and vice versa. Importantly, a licensed shift allowed one meaning to be generated from the other word, but not the other way around. At the same time, we gathered information about the similarity of the two words' meanings, whether one meaning was conceptually central for the other and whether they were related through either cue validity or salience. Current radical pragmatic theories predict that sense acceptability should be tightly related to at least one of our conceptual measures. In particular, most theories predict that acceptability should increase as conceptual centrality increases.³ By contrast, fully rule-based theories predict that senses within the semantic field of a licensed polysemy rule should be acceptable and make no predictions about the relationship between sense acceptability and conceptual structure.

Study 2 went beyond a regression-based analysis. We constructed a number of artificial vocabulary items (as in Murphy 1997), which varied both in whether they could be entered into a licensed shift and in whether the possible shifted meaning constituted a central part of the artificial item's description. This allowed us to test the causal efficacy of these two factors in creating a shiftable meaning.

2 STUDY 1

2.1 *Methods*

2.1.1 *Participants* Forty-six participants provided judgments of sense acceptability via the paraphrase task. A different set of 12 participants provided judgments for each conceptual metric. All were college undergraduates receiving course credit for participation.

³ Note that other possible radical pragmatic theories could potentially rely on metrics that are untested here, so it will not be possible to definitively rule out the whole class of theories based on these data.

2.1.2 *Materials* We took 60 pairs of words (e.g. *DVD*, *movie*), 15 for each of four known regular polysemy patterns, which changed the meaning of one word to the meaning of its pair (i.e. the word *DVD* has a sense akin to the word *movie*) but not vice versa. In the following discussion, reference to an item describes an ordered word pair [e.g. the pairs (*DVD*, *movie*) and (*movie*, *DVD*) are different items].

The four regularities used were object–abstract object (e.g. *DVD*, *movie*), container–contents (e.g. *pot*, *liquid*), producer–product (e.g. *writer*, *book*) and institution–person (e.g. *hospital*, *doctor*). We then devised an additional 120 sentence frames that acted as contexts to select for the senses (e.g. *was an hour long*, *was round*). We chose word pairs based on our intuition that they fell into the relevant semantic fields (e.g. containers) for the licensed and unlicensed rules. We chose contexts based on our intuition that they strongly selected for the default sense of a word (i.e. the abstract meaning of *movie*, the object meaning of *DVD*).

The word pairs were matched in number of morphemes and length. However, we were unable to control for frequency: the words that did not enter the regular shift had higher frequency. Given that more frequent words are in fact known to be more polysemous (Zipf 1945), we assumed that this difference should not affect our results, but we also carried out an analysis that accounted for it (see footnote 5 in section 2.2).

2.1.3 *Measuring sense acceptability* How do you assess whether a word like *DVD* has both a disc and a film sense? One possibility would be to collect informant judgments (as for syntactic acceptability), but this requires considerable metalinguistic contemplation, and informants may have difficulty recalling particularly low-frequency senses. Instead, we settled on paraphrase judgments, in which we tested whether the target word (*DVD*) could acceptably replace its paired word (*movie*). Since the target and paired words had very different default meanings (*DVD* = disc, *movie* = film), a paraphrase was only acceptable if the sense of the target could be shifted to something like the meaning of its paired word. For example, participants judged whether they agreed or disagreed with statements like “The movie was an hour long” can be paraphrased as “The DVD was an hour long” or “The DVD was shiny” can be paraphrased as “The movie was shiny”.

Participants made these judgments by sliding a bar on a continuous scale between the poles *disagree* and *agree*, and ratings were *z*-scored within subjects. All stimuli were presented using scripts developed using the PyPsyExp Python library (Gureckis 2009). We used a Latin

square design, such that each participant received only one of the two possible items from each pair. Items are given in Appendix A.

2.1.4 Measuring the relatedness of meanings We estimated the relevant conceptual metrics by collecting verbal ratings from participants. We chose these metacognitive tasks over other behavioural measures such as priming or classification because it is unclear exactly which lexical or conceptual constructs are tapped by behavioural tasks. To measure similarity, we simply asked participants to make judgments of the form ‘How similar is a DVD to a movie?’ on a continuous scale between *not at all similar* to *very similar*. Before answering, participants were told that the questions were supposed to measure the extent to which two items were the same sort of thing and were given examples of both similar and dissimilar pairs of items.

Calculating centrality was more complicated. As a measure of conceptual centrality, we used a metric from Sloman *et al.* (1998): judgments of similarity to an ideal. In particular, we asked participants to rate how much the absence of a relationship with the to-be-shifted sense made the target’s meaning less ideal. All questions were of the same form: ‘Think of an ideal DVD. How similar is a DVD that is not associated with a movie to an ideal DVD?’ If a sense is particularly central in the target word’s concept, then other features should also depend on it, and so its absence should make the concept considerably less ideal. Participants provided ratings on the scale *not at all similar* to *very similar*. We inverted these ratings to create a centrality score (low similarity to ideal = high centrality). Before answering the questions, participants were instructed to think about how they would describe an ideal version of an example object (e.g. a mechanic). Then, they were instructed to imagine that object, but now not associated with another object (e.g. a mechanic not associated with tools). Finally, they were asked whether that new object was similar to the original ideal version.

When measuring centrality, we used ‘associated with’ rather than provide a specific relation (e.g. *contains a movie*). This was to avoid conflating the centrality of a sense with the specified relation. We hoped to abstract over different cases, and thereby collect the average centrality of a concept.

We also used Sloman *et al.*’s (1998) methods to measure cue validity and salience. Again, all questions took the same basic form. For cue validity, participants answered questions such as ‘Of all the things associated with a DVD, what percentage are movies?’ (rating the answer from 0% to 100%). If the concept DVD is a valid cue to the concept movie, the latter should be frequently associated with the former. To

some extent, this question also measures diagnosticity, which describes the extent to which *DVD* cues *movie* and only *movie*. For instance, if *DVD* validly cues *movie*, as well as many other concepts (i.e. it is not particularly diagnostic), participants should provide a low rating. We considered this conflation acceptable because prior theories relying on cue validity did not consider the importance of diagnosticity, and yet, diagnosticity *a priori* seems just as likely to play a role as cue validity (if the measure provides a handle on sense acceptability, follow-up work could distinguish the two possible causes.) To ensure participants were answering the questions as intended, they were instructed to think about things associated with some examples (e.g. associates of a mechanic), and what percentage of those things were a second type of thing (e.g. some tools).

Finally, to assess the relative salience of the two concepts, participants made judgments of prominence: ‘When you think about a DVD, to what extent do you think about a movie?’ (rating from *not at all prominent* to *very prominent*). Salient concepts should frequently come to mind and have a prominent place in participants’ thoughts, even when they are neither cue valid nor central. Before answering, participants were instructed to think about whether, when thinking about one example item (e.g. a mechanic), another item (some tools) would be one of the first things to ‘pop into their head’.

Judgments were again made on a continuous scale, with ratings z -scored within individual subjects. Again, we used a Latin square design, with each participant receiving one of the two possible items per pair. Examples of each question type are given in Table 2 below.

2.1.5 Procedure Participants were tested, up to four at a time, at individual computer terminals. Prior to each test, participants were given a training period. For the paraphrase task, the training was entirely computer based. Participants were told they would be rating the extent to which one sentence conveyed the same meaning as another despite the difference in words and were shown a set of example paraphrases (e.g. *Mary was happy* cannot be paraphrased as *The tree was happy*, *John was happy* can be paraphrased as *The man was happy*). Finally, they were shown how to give ratings using the slider.

Participants in the conceptual tasks were told that the experiment was designed to elicit their ratings of how everyday things are related to other everyday things. Training began with a short discussion by the experimenter on the aims for the particular measure and description of how to think about two examples. An example script is given in Appendix B. The relevant descriptions are given in section 2.1.4.

Metric	Question
Similarity	How similar is a DVD to a movie?
Centrality	Think of an ideal DVD. How similar is a DVD that is not associated with a movie to an ideal DVD?
Cue validity	Of all the things associated with a DVD, what percentage are movies?
Saliency	When you think about a DVD, to what extent do you think about a movie?

Table 2 Examples of the questions asked for each metric for the (*DVD*, *movie*) pair [the names were reversed for the pair (*movie*, *DVD*)]

After this discussion, participants were walked through two practice questions on the computer, in which they were shown how to answer using the slider. They were told the experimenter's intuition about what the answer to each question should be, but it was emphasized that their judgments in the experiment should reflect their own intuitions. Finally, the experimenter elicited and answered questions.

2.2 Results

Figure 1 shows that the sense acceptability ratings were reliably greater for those items that entered into a licensed regularity [licensed: mean = 0.53, standard deviation (SD) = 0.56; unlicensed: mean = -0.53, SD = 0.45; $t(118) = 11.45$, $P < 0.01$], but there was still considerable variability within the two sets, which overlapped. We next assessed what best predicted that variation. If the sole contributor to sense acceptability is a set of rules, we would not expect to find that any of the conceptual metrics predict sense acceptability. But if any metric does come out as a strong predictor, that would lend support to a radical pragmatics-type theory.

To assess these hypotheses, we simultaneously regressed sense acceptability against each of our conceptual metrics, using a multilevel model with random intercepts (Gelman & Hill 2007) for the different regular polysemy patterns ($n = 120$, deviance = 238).⁴ In addition, Table 3 shows the first-order correlations between the various conceptual metrics and sense acceptability.

Following Klein & Murphy (2002: 556), who found that extant polysemous senses were generally not rated as similar, we found that the degree of similarity between a word's meaning and a potential sense did not predict whether that sense was acceptable. Figure 2 shows that there was no major difference in similarity scores between the items

⁴ We calculated P values using Markov Chain Monte Carlo estimates of the coefficients, using the `pvals.fnc` function in the *languageR* library (Baayen 2008).



Figure 1 Points show normalized sense acceptability ratings for each item, split by participation in a licensed regular shift (black circles, e.g. the word *DVD* used with a movie sense) or an unlicensed shift (grey squares, e.g. the word *movie* used with a DVD sense). Thick black lines indicate mean score; * $P < 0.05$.

	Similarity	Centrality	Cue validity	Salience
Sense acceptability	0.08	0.33	0.31	0.23
Similarity		0.45	0.50	0.50
Centrality			0.74	0.74
Cue validity				0.77

Table 3 First-order correlations between the different measures used

that did and did not fall into a licensed shift (licensed: mean = -0.02 , SD = 0.66 ; unlicensed: mean = 0.02 , SD = 0.61 ; $t(118) = 0.33$, *ns*), and more importantly, similarity did not reliably predict sense acceptability [$\beta = -0.16$, standard error (SE) = 0.12 , $t = 1.3$, *ns*].⁵ This is not consistent with any radical pragmatics hypothesis that takes similarity to be the major conceptual metric determining polysemy.

Centrality, our operationalization of noteworthiness, proved to be a more reliable predictor of sense acceptability: the extent to which a potential sense was central for the initial meaning was related to sense acceptability. However, the effect was complicated. Figure 3a shows that there was a small difference in centrality between senses that fell into licensed and unlicensed shifts, with the latter less likely to be central (licensed: mean = 0.28 , SD = 0.77 ; unlicensed: mean = -0.28 ,

⁵ This result, and the results for centrality, cue validity and salience, did not change when lexical frequency of the items was included in the regression.

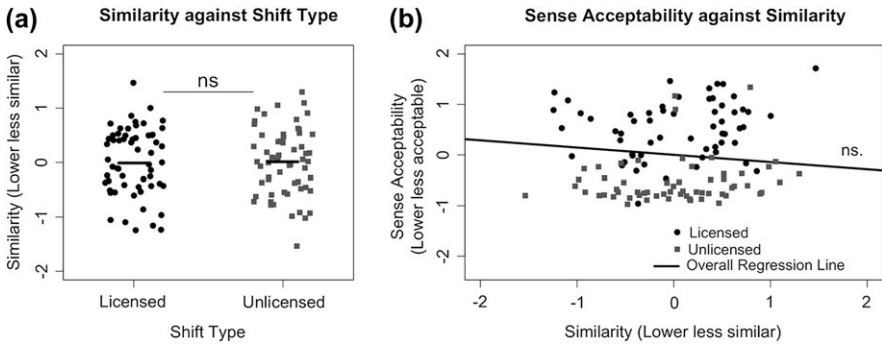


Figure 2 (a) Difference in similarity scores between word pairs related via a licensed or an unlicensed shift. (b) Sense acceptability score of a word pair against similarity score.

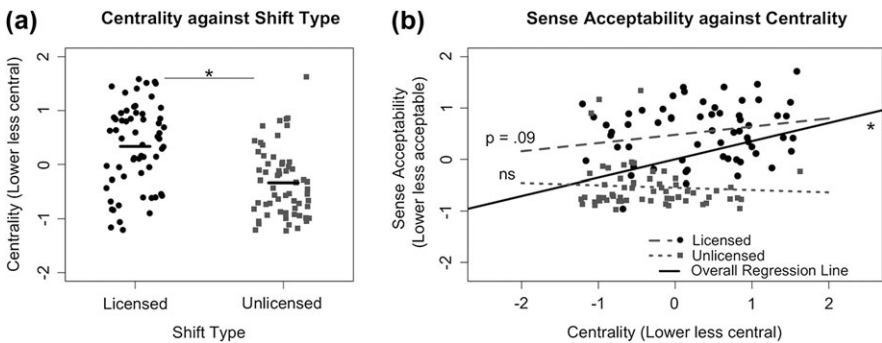


Figure 3 (a) Difference in centrality scores between word pairs related via a licensed or an unlicensed shift. (b) Sense acceptability score of a word pair against centrality score.

SD = 0.64; $t(118) = 4.4$, $P < 0.01$). Furthermore, sense acceptability increased with increasing centrality ($\beta = 0.33$, SE = 0.14, $t = 2.39$, $P = 0.02$; Figure 3b).

But direct inspection of the second graph in Figure 3 suggests that it is not simply the case that being central automatically makes a sense acceptable. In particular, note that the points in Figure 3b appear to fall into two clusters, senses that are derived by a licensed shift (black circles) and senses that are not (grey squares). Between these clusters, there is a large reliable difference in sense acceptability but only a smaller difference in associated centrality scores: compared with senses derived by an unlicensed shift, senses derived by a licensed shift tend to describe concepts that play a more central role in the default senses. However, within each cluster, centrality is not a good predictor, despite its variability. The dashed lines in Figure 3b show that the effect of centrality on acceptability was not reliable for the unlicensed items ($\beta = -0.03$, SE = 0.11, $t = 0.14$, *ns*) and

was only marginal for the licensed items ($\beta = 0.18$, $SE = 0.1$, $t = 1.78$, $P = 0.09$). This would not be expected under a centrality-based gradient account. It is more consistent with a rule-based model, in which falling under a rule is the primary determinant of sense acceptability and centrality plays a limited role at best.

Finally, we assessed whether our other conceptual measures reliably predicted any additional variance in sense acceptability. They did not (cue validity: $\beta = 0.16$, $SE = 0.13$, $t = 1.2$, *ns*; salience: $\beta = 0.06$, $SE = 0.13$, $t = 0.47$, *ns*), although as can be seen in Table 3, the two measures were correlated with centrality, suggesting that a common factor may contribute to variance in each. Still, because conceptual centrality offered the most predictive power, a provisional conclusion should be that, to the extent polysemy rules are in any way dependent on world knowledge, the metric they depend on is similar to centrality.

2.3 Discussion

Study 1 tested whether the acceptability of extended senses for a large set of words could be explained by how related they were to their default meaning, or whether a more traditional rule-based approach provided a better fit to the data. We demonstrated that one conceptual metric, similarity, was a very poor predictor of sense acceptability, despite its importance for theories of concepts and categories. This replicated previous work by Klein & Murphy (2001, 2002) and extended it to additional forms of regular polysemy. In tandem with Klein and Murphy's results, we think the present data provide relatively conclusive evidence that the extent to which two concepts overlap in meaning has very little effect on sense acceptability.

The best indication of an effect of conceptual relatedness came from our centrality measure: acceptable senses tended to be more central, and cue validity and salience did not predict any additional variation in acceptability. This is consistent with theories in which an intrinsic relation between concepts licenses polysemy (e.g. Fauconnier 1985; Nunberg 1995; Wilson 2003), but not with the theories of Nunberg (1979) and Papafragou (1996), which specifically rely on cue validity and salience, respectively.

However, the effect of centrality was not particularly robust: within the set of unlicensed shifts, centrality did not reliably predict acceptability, and within the set of licensed shifts, it only marginally predicted acceptability. Instead, the effect occurred because items participating in a licensed shift tended to be considerably more acceptable overall but only a little more central.

What could explain the small but reliable effect of conceptual centrality? It may be that both radical pragmatics and rule-based models are correct, and centrality and polysemy rules act relatively independently to produce word meanings. Copestake & Briscoe (1995) suggest such a two-system account, in which both polysemy rules and more pragmatic transfers of reference can occur. But this would not predict centrality's lack of an effect within the licensed and unlicensed clusters, unless it was simply an issue of power.

An alternative is that centrality predicts the occurrence of polysemy rules: rules tend to describe relationships that are central for a large number of lexical items. However, once the rule is in place, its application is more or less independent of centrality. This would account for the small difference in centrality between licensed and unlicensed senses.

A final possibility is that polysemy rules rely, to an extent, on centrality: for a rule to operate properly on a word, that relationship has to be relatively central. For instance, the word *pot* can be entered into a container–contents rule because the fact that pots are containers is a relatively central aspect of the concept underlying the word. By contrast, the word *tree* cannot be entered into the same rule because trees are not obvious containers. This would explain the marginal effect of centrality on sense acceptability for items from the licensed shifts and the nonreliable effect for items from the unlicensed shifts.

In summary, we see three possible explanations for the centrality results in Study 1.

- (i) Centrality and polysemy rules both cause sense acceptability, and we lacked the power to determine this.
- (ii) Centrality does not cause sense acceptability but does predict the presence or absence of a polysemy rule.
- (iii) Polysemy rules can only be used when the relationship they describe is central for the concept.

3 STUDY 2

To evaluate these hypotheses, we created a set of artificial lexical items and tested how easily they could be shifted to a different sense. This allowed us to experimentally manipulate whether the meaning of a word had a central relationship to another possible sense and whether the shift in sense fell under a licensed template. By crossing these factors, we could tease apart the true determinants of sense acceptability. It also allowed us to confirm that the effects in Study 1 reflected causation rather than correlation because subjects here had no experience with the secondary sense prior to our collecting their judgment.

Half our artificial words described novel types of containers, which could enter into a licensed container–contents shift. The other half were potential contents, in particular, different types of substances. These could only enter into an unlicensed contents–container shift. To make the shifted meaning central, we designed coherent reasons why it would be important for the item. For example, one container (a ternway, see Example i) was designed to hold water, and one example of some contents (some zav, see Example ii) had to be kept in a special container because it was a powerful solvent. To make the shifted meanings non-central, we simply mentioned them as incidental parts of the concept: the ternway happened to have water in it or the zav just happened to be in a particular container.

We used a similar paraphrase task to measure sense acceptability for the novel items but embedded in slightly more context in order to provide more information. Participants first read a description of the novel word’s meaning, in which centrality was varied. This was followed by a short scenario in which the item was used.

The final sentence of the scenario contained a reference to the intended shifted meaning of the item [in Example i, the contents of the ternway (water), and in ii, the container of the zav (a jar)]. Participants were asked to judge whether that sentence was well paraphrased if the novel word was substituted in place of the alternative meaning. Ratings were made on a scale between 1 (*unacceptable to switch*) and 7 (*totally acceptable to switch*). The same scenarios were used for both the central and non-central versions of each item.

(i) *Description:* A ternway is a large metal object. Ternways are found on farms. They are strong but hollow.

Central: Farmers fill them with water, so that animals can drink from them.

Non-central: Farmers use them to build temporary constructions and shelters for their animals.

Scenario: One day, Harry was inspecting one of his ternways. It was full of water. At that moment, his neighbour, Bob walked by. Bob stopped to say hello. As they chatted, Harry rested his hand on the ternway. **He absent-mindedly stirred the water with his finger.**

Question: How acceptable would it have been to instead have said that ‘He absent-mindedly stirred the ternway with his finger’ and have meant the same thing?

(ii) *Description:* Zav is a watery substance. It is produced by cactuses.

Central: Zav is always kept in a metal can because it corrodes any other materials.

Non-central: Zav is very corrosive.

Scenario: Zav is used to clean off stains. Harry and Jane have just purchased an old, tumbledown house. They plan on cleaning and fixing it up. Harry goes to work cleaning the second floor of the house. He carries a lot of equipment upstairs with him, and on the way up, he drops the can of zav, which crashes down. Jane rushes over to see what the noise was. Harry calls down to see if everything is ok. Jane answers that everything is fine and that the only damage was that ‘**The can is dented**’.

Question: How acceptable would it have been for Jane to instead have said, ‘The zav is dented’ and have meant the same thing?

3.1 Methods

3.1.1 *Participants* Thirty-two college-age students participated, in exchange for course credit.

3.1.2 *Materials* We constructed six items, three containers and three contents, each with a central and a non-central variant, as described above, and listed in Appendix C.

3.1. *Procedure* Participants read and answered each item using paper and pencil. In order to control for each participant’s prior willingness to shift, as well as to instruct them on the task, they first answered four scenarios using extant licensed and unlicensed English shifts (such as from *DVD* to *movie* and vice versa). We normalized acceptability judgments for the target items over the mean of these scores.

Following these instructions, each participant received six novel words, half in their central form, half non-central, in one of three different orders.

3.2 Results

If centrality and polysemy rules contribute to sense acceptability independently, both effects should be reliable. If licensed shifts make senses acceptable, we only expect an effect of shift type. Finally, if centrality affects the application of licensed polysemy rules, the two factors should interact: while centrality should not influence the acceptability of unlicensed items, higher centrality should increase the acceptability of a licensed item.

Sense acceptability ratings (normalized over mean practice item ratings) are displayed in Figure 4. We analysed them using a 2×2

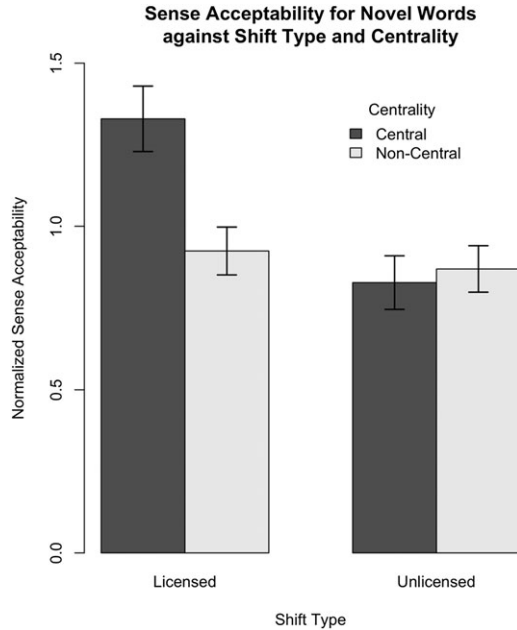


Figure 4 Sense acceptability against shift type and centrality. Acceptability is calculated as score assigned to the paraphrase normalized over the mean of the scores assigned to the practice items by the participant.

repeated measures analysis of variance (shift type: licensed/unlicensed; centrality: central/non-central), which revealed a reliable main effect of shift type [licensed: mean = 1.13, SD = 0.55; unlicensed: mean = 0.85, SD = 0.41; $F(1,31) = 13.19$, $P < 0.01$] and a main effect of centrality [central: mean = 1.07, SD = 0.55; non-central: mean = 0.89, SD = 0.43; $F(1,31) = 9.99$, $P < 0.01$]. But importantly, both effects were qualified by a reliable interaction between shift type and centrality: $F(1,31) = 6.7$, $P = 0.01$. In particular, while centrality had a reliable effect on sense acceptability for items falling under a licensed shift, $t(31) = 4.61$, $P < 0.01$, it did not have a reliable effect for items falling under an unlicensed shift, $t(31) = -0.36$, *ns*. This suggests that centrality is not an independent contributor to sense acceptability. Instead, senses are produced by the application of rules, and their acceptability is determined by their centrality.

Finally, we examined whether simply being a potential container was enough to increase sense acceptability, regardless of whether that facet of the concept was made central. In fact, the non-central items that entered a licensed shift were no more acceptable than either the

central or non-central items that entered an unlicensed shift (both $t_s < 1.3$, *ns*). This, again, is consistent with the proposal that, to enter a shift, the relevant part of the concept has to be made central.

4 GENERAL DISCUSSION

Although word meanings are flexible, they cannot change in infinite ways. A DVD can be shiny, or a DVD can be an hour long, but while a movie can be an hour long, it cannot be shiny. Rule-based models assume that these limits occur because a finite number of shifts generate possible senses. By contrast, radical pragmatic theories propose that the limits to word meaning reflect some structure in our concepts, so that the acceptability of a derived sense depends on whether it is closely related to the default sense.

In our data, the acceptability of a sense was well predicted by whether a word could enter into a polysemy rule but, *contra* the radical pragmatics models, it was not well predicted by any conceptual metrics. Conceptual centrality, which we argued was critical for the theories of Nunberg (1995), Fauconnier (1985) and various versions of relevance theory (Papafragou 1996; Wilson 2003), offered some predictive power but not independently of proposed lexical rules. Other metrics, such as similarity, cue validity and salience offered no explanatory value above and beyond that provided by centrality. The results therefore provide evidence against the radical pragmatic accounts of sense shifts and in favour of an account that relies on conventional rules.

As discussed in the introduction, various radical pragmatic theories have also endorsed the possibility that there might be polysemy conventions that exist alongside, or are built on top of, the pragmatic shifting system (Nunberg 1995; Papafragou 1996). These types of rules and conventions are consistent with our data, but the theories proposing them also necessarily predict that relationships between concepts should explain sense acceptability independently of any lexical rule (Fauconnier 1985; Nunberg 1979, 1995; Papafragou 1996). This prediction was not met, for instance, centrality did not affect acceptability for the unlicensed contents–container shifts in Experiment 2.

Our results are therefore more consistent with rule-based models than radical pragmatic ones. But while a fully gradient effect did not emerge in our data, Study 1 did demonstrate that senses produced by a rule have a tendency to be more central than senses that did not fall under a licensed rule. And moreover, senses produced by rules tended to be more acceptable when they had a central, or intrinsically important, connection to the original meaning. These results are not

consistent with the radical pragmatics models discussed in the introduction, but nor are they consistent with a version of the rules model in which there is no link between concepts and the rules themselves, for example, a model in which rules are just arbitrary features of a language (e.g. Ostler & Atkins 1992).

Instead, the results support a middle ground, partially built on the conventionalization accounts discussed above. In particular, we propose that (a) aspects of world knowledge such as centrality do not directly cause sense acceptability, but (b) conventional polysemy rules, which do license senses, arise to mark generally important and central relationships. Finally, (c) the acceptability of the output of a polysemy rule (e.g. container–contents) is determined by whether the affected word meaning is a prototypical example of that relationship (i.e. a good example of a container). Importantly, this latter factor will be partially determined by centrality. For example, a good example of a container will be one whose contents play a central role in how it is conceptualized.

There are certain caveats for this model given the present data. First, our claim is based on a set of null results, which means that we cannot rule out whether another factor might predict acceptability, such as a differently defined version of relevance. An effect might also be found using different means of collecting the same relations, either through alternately worded questions or a very well defined implicit task such as sorting or priming. But while it is risky to make broad inferences from such data, we believe that the large number of tested conceptual relations justifies it.

Second, we only had limited control over the effects of context in the sense acceptability judgments of Study 1. Context modulates the acceptability of a sense, but our acceptability judgments only used one context per sense, and the strength of this context was only set via intuition. It may be that some of the extreme acceptability judgments from Experiment 1 resulted from poorly chosen contexts (for instance, adjectives that were polysemous⁶). Still, we suspect that our overall claims are robust because of the high number of tested items.

Finally, there is one situation where centrality might still have a direct effect on interpretation: deciphering fully novel references, as in Nunberg's (1995) example *The ham sandwich is at Table 7* (spoken in reference to a customer by a restaurant server). In these cases, an interpretation of *ham sandwich* as a food item is implausible, and so the speaker forces the listener

⁶ For instance, one question asked participants if 'The book was thin' could be paraphrased by 'The story was thin'. The acceptability clearly depends on whether thin is assigned a size sense or the less frequent poor quality/insufficient sense.

to adopt a new interpretation. The search for an appropriate novel meaning might rely on centrality as a guide, both in comprehension and in production (cf. Clark & Gerrig 1983). An important goal for future research is to understand exactly what guides the resolution of novel deferred references and to understand when speakers might choose to force listeners to comprehend this type of fully novel reference. Conceptual structure may prove important in both cases.

The current results have particularly interesting implications for both the occurrence of regular polysemy cross-linguistically and the development of polysemy rules in children. If polysemy rules are conventions, we expect considerable cross-linguistic variation in their occurrence. Although cross-linguistic polysemy has not been systematically surveyed, isolated reports provide at least some support for this prediction. For instance, according to Kamei & Wakao (1992), the producer–product rule (for which they tested the example *he read Mao*) is considerably more acceptable in English and Japanese than in Mandarin Chinese.

Conventional polysemy rules also set up a tricky learning problem for children, who need to avoid over- or underestimating the licensed patterns. This requires at least some restrictions on possible rules; otherwise, children would initially consider far too broad a range (e.g. they might search for evidence for or against a headwear–crustacean rule). A metric such as centrality might provide that restriction: children could set an upper bound on possible rules by assuming that polysemy only encodes particularly central relationships (see Schütze 1997 for a similar proposal regarding subcategorization frames). Rabagliati *et al.* (2010) provide some tentative evidence for this. They show that relatively young children (4–6 years) overestimate which senses are possible, but they do not wildly overestimate. While they accept licensed senses (e.g. *the hour-long DVD*) and some unlicensed senses (e.g. *the shiny movie*), they reject senses that cannot be easily connected to the original meaning (e.g. *the angry rock*). This pattern of early generality and later rule-based specificity suggests considerable developmental change in lexical representations: while adults assign senses based on rules, children may well be radically pragmatic.

Acknowledgements

We would particularly like to thank Gregory Murphy for a number of helpful discussions during this project. Additional thanks are due to Sandeep Prasada, Marjorie Rhodes, Josh Hartshorne, Manizeh Khan, Anja Sautmann, Jesse Snedeker, Mahesh Srinivasan and

Caran Wakefield, as well as audiences at NYU, Brown, Harvard and Yale. This work was supported by National Institutes of Health grant R01-HD48733.

APPENDIX A

A. Experiment 1 stimuli. Each bracket contains (Noun1 [enters unlicensed shift], Context1 [selects Noun1 sense] & Noun2 [enters licensed shift], Context2 [selects Noun2 sense]).

Container–content

(liquid, stirred & pot, cracked); (water, boiling & kettle, smashed); (juice, poured & pitcher, broken); (magic potion, cooled & cauldron, metal); (milk, spilled & glass, plastic); (drink, delicious & bottle, warped); (flowers, fragrant & vase, chipped); (meal, filling & plate, glazed); (tea, steaming & mug, porcelain); (tuna, expired & can, dented); (jam, fresh & jar, dusted); (food, tasty & bowl, split); (meat, sizzling & pan, scratched); (trash, recycled & dumpster, graffitied); (sand, leaking & bucket, leaky)

Institution–person

(pilot, contacted the tower & plane, twin-engined); (shopkeeper, wrote to the customer & store, made by a builder); (doctor, called the lady & hospital, built out of bricks); (receptionist, answered the phone & hotel, four stories tall); (students, appeared on TV & kindergarten, contained 10 classrooms); (tour, greeted the visitors & museum, guide closed in the winter); (mayor, hall popular & city, baroque); (monarch, criticized & palace, whitewashed); (monks, sang beautifully & monastery, beautifully decorated); (teacher, listened to the parents & school, designed in New York); (motorist, spoke with the security guard & car, repaired); (driver, waved at the group & truck, weighed two tons); (people, liberal & city, 7 miles wide); (passengers, talkative & train, had four carriages); (captain, fired the crew & ship, moored)

Object–abstract

(story, about a boy & book, thin); (adventure, set in space & comic, red and green); (movie, an hour long & DVD, round); (song, loud & CD, shiny); (show, about animals & videotape, plastic); (plot, well paced & novel, in hardback); (recording, poorly mastered & cassette, snapped); (music, explicit & mp3, zipped); (information, well researched &

encyclopaedia, 100lb); (news, shocking & newspaper, torn); (food, delicious & recipe, well-written); (message, urgent & note, folded); (slogan, rude & t-shirt, size 5); (pattern, symmetrical & flag, torn); (documents, controversial & binder, shut)

Producer–product

(book, proof-read & writer, bald); (music, performed & composer, smelly); (building, commissioned & architect, laughing); (sculpture, archived & sculptor, rich); (novel, re-issued & novelist, tall); (picture, displayed & artist, smiling); (song, recorded & singer, trained); (painting, catalogued & painter, hairy); (film, edited & filmmaker, rich); (article, published & journalist, unwashed); (play, rediscovered & playwright, shy); (photograph, exhibited & photographer, loud); (dance, technical & dancer, well-trained); (poem, read & poet, washed); (design, bootlegged & designer, arrogant)

APPENDIX B

A. Example training script (centrality)

We are going to ask you to do some tasks. In the first task, we will ask you questions about a number of different everyday things and people, and how they relate to other everyday things and people. In particular, we're going to ask you to imagine an ideal version of an object, say, some ideal fruit. Then, we're going to ask you about that same object that's missing something it may be normally associated with. We want you to tell us whether missing that thing affects whether that object is ideal or not.

This is a little complicated, so let's go through an example. First, think about fruit. Think about what some ideal fruit would be like. Now, imagine that you found some fruit that was not associated with a fruit vendor on the street. How similar would that fruit be to some ideal fruit? To what extent does 'not being associated with a fruit vendor' affect whether something is ideal?

Probably, your answer is that it is not important and that fruit is still relatively similar to some ideal fruit. For example, it could be plucked straight off a tree and eaten, and that seems like a pretty ideal piece of fruit.

Here's another question: Think about an ideal mechanic, the sort of mechanic you would imagine if you had to describe what you would want a mechanic to be like. Now, how similar is a mechanic who is not

associated with any tools to an ideal mechanic? This one is harder. It's pretty important to being a mechanic that you use tools a lot, so a mechanic who never uses tools would be non-ideal. So here, you'd probably say that the mechanic is not very similar to an ideal mechanic. We're going to ask you these questions, and you will answer them, via the computer terminals in front of you [participants then answer practice questions on the computer terminal].

APPENDIX C

A. Items used in Experiment 2

Container–contents

1. **Description:** A blittle is a spherical, plastic object with a flat base and a hole in the top.

Central: Blittles are made in Ecuador, where they are used to store marbles. Their round shape means that they can store a lot of marbles.

Non-central: Blittles are made in Ecuador, where they are given as gifts.

Scenario: Maria was playing with her son Juan. Suddenly, the phone rang. It is an important call, and Maria needs to talk for a long time, so she gave Juan a blittle, which she had filled with marbles, to play with. After the phone call, Maria finds Juan. She sees that he has taken all the marbles and put them into groups of the same colour. She asks him what he did, and he replies, '**I sorted the marbles**'.

Question: How acceptable would it have been for Juan to instead have said, 'I sorted the blittle' and have meant the same thing?

2. **Description:** A ternway is a large metal object. Ternways are found on farms.

Central: They are strong but hollow. Farmers fill them with water, so that animals can drink from them.

Non-central: They are strong but hollow. Farmers use them to build temporary constructions and shelters for their animals.

Scenario: One day, Harry was inspecting one of his ternways. It was full of water. At that moment, his neighbour, Bob walked by. Bob stopped to say hello. As they chatted, Harry rested his hand on the ternway. **He absent-mindedly stirred the water with his finger.**

Question: How acceptable would it have been to instead have said that, 'He absent-mindedly stirred the ternway with his finger' and have meant the same thing?

3. **Description:** A shiffle is a large wooden thing. It has pretty engravings on the outside, and is hollow.

Central: Shiffles are made in New Zealand, where they are used to collect and store fruit.

Non-central: Shiffles are made in New Zealand, where they are used for decoration in people's gardens.

Scenario: John owns a ranch on New Zealand's South Island. Behind his house, there is a beautiful orchard, filled with fruit trees. John had harvested a lot of apples, and left them in his shiffle. Unfortunately, the apples started to rot. When John found out, he was very disappointed. **He emptied the apples into the compost near the orchard.**

Question: How acceptable would it have been to instead have said that 'He emptied the shiffle into the compost near the orchard' and have meant the same thing?

Contents–container

4. **Description:** Backle is a gooey black candy made out of sugar.

Central: Backle is kept in small plastic packets so you can take it hiking, because it's a great energy source.

Non-central: Backle is full of energy, so that people eat it when they go hiking.

Scenario: Jane and Larry were planning on going for a long hike. Jane went to the store and bought some backle. Later, on their hike, Larry and Jane are walking through some wild roses. Unfortunately, Larry catches his packet of backle on a thorn, and it tears. He curses, and tells Jane **'I caught my packet on a thorn'**.

Question: How acceptable would it have been for Larry to instead have said 'I caught my backle on a thorn' and have meant the same thing?

5. **Description:** Blika is a soft, gooey and green substance. It grows on trees in Central Alegre, in Brazil.

Central: Blika is always kept in special jars, so that you can heat it, because when it is hot it drives away insects.

Non-central: When you heat blika, it drives away insects.

Scenario: Ricardo and Maria are having a small party. Unfortunately, there are lots of mosquitoes around that evening, and they don't have any blika left in their house Ricardo goes to the store to buy some. Unfortunately, he slips and falls on the way home, and the jar of blika gets shattered. When he gets home he apologizes to Maria, saying **'I broke the pot!'**

Question: How acceptable would it have been for Ricardo to instead have said ‘I broke the blika!’ And have meant the same thing?

6. **Description:** Zav is a watery substance. It is produced by cactuses.

Central: Zav is always kept in a metal can because it corrodes any other materials.

Non-central: Zav is very corrosive.

Scenario: Zav is used to clean off stains. Harry and Jane have just purchased an old, tumbledown house. They plan on cleaning and fixing it up. Harry goes to work cleaning the second floor of the house. He carries a lot of equipment upstairs with him, and on the way up, he drops the can of zav, which crashes down. Jane rushes over to see what the noise was. Harry calls down to see if everything is ok. Jane answers that everything is fine and that the only damage was that ‘**The can is dented**’.

Question: How acceptable would it have been for Jane to instead have said, ‘The zav is dented’ and have meant the same thing?

Hugh Rabagliati

Department of Cognitive, Linguistic and Psychological Sciences

Brown University

Providence, RI 02912

USA

e-mail: hugh_rabagliati@brown.edu

REFERENCES

- Ahn, W. K. (1998), ‘Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality’. *Cognition* **69**:135–78.
- Ahn, W. K., S. A. Gelman, J. A. Amsterlaw, J. Hohenstein & C. W. Kalish. (2000), ‘Causal status effect in children’s categorization’. *Cognition* **76**:B35–43.
- Apresjan, J. (1974), ‘Regular polysemy’. *Linguistics* **142**:5–32.
- Baayen, R. H. (2008), *languageR: Data Sets and Functions with Analyzing Linguistic Data: A Practical Introduction to Statistics*. R package version 0.953. University of Alberta, Alberta, CA.
- Borer, H. (2005), *In Name Only*. Oxford University Press. Oxford.
- Clark, H. H. & R. J. Gerrig. (1983), ‘Understanding old words with new meanings’. *Journal of Verbal Learning and Verbal Behavior* **22**:591–608.
- Copetake, A. & E. J. Briscoe (1995), ‘Semi-productive polysemy and sense extension’. *Journal of Semantics* **12**:15–67.
- Dölling, J. (1995), ‘Ontological domains, semantic sorts, and systematic ambiguity’. *International Journal of Human-Computer Studies* **43**:785–807.

- Fauconnier, G. (1985), *Mental Spaces*. MIT Press. Cambridge, MA.
- Gelman, A. & J. Hill (2007), *Data Analysis Using Regression and Multi-level/Hierarchical Models*. Cambridge University Press. New York.
- Gibbs, R., D. Beitel, M. Harrington & P. Sanders. (1994), 'Taking a stand on the meanings of stand: bodily experience as motivation for polysemy'. *Journal of Semantics*, **11**:231.
- Gureckis, T. M. (2009), *PyPsyExp System*. New York University. New York (<http://smash.psych.nyu.edu/pypsyexp/>). Last accessed: 04/06/2011.
- Kamei, S.-i. & T. Wakao (1992), Metonymy: reassessment, survey of acceptability, and its treatment in a machine translation system. *Paper presented at the Proceedings of the 30th annual meeting of the Association for Computational Linguistics*.
- Klein, D. E. & G. L. Murphy (2001), 'The representation of polysemous words'. *Journal of Memory and Language* **45**:259–82.
- Klein, D. E. & G. L. Murphy (2002), 'Paper has been my ruin: conceptual relations of polysemous senses'. *Journal of Memory and Language* **47**:548–70.
- Murphy, G. L. (1997), Polysemy and the creation of novel word meanings'. In T.B. Ward, S.M. Smith, and J. Vaid (eds.), *Creative Thought: An Investigation of Conceptual Structures and Processes*. APA. Washington, DC. 235–65.
- Murphy, G. L. & D. L. Medin. (1985), 'The role of theories in conceptual coherence'. *Psychological Review* **92**:289–316.
- Nunberg, G. (1979), 'The non-uniqueness of semantic solutions: polysemy'. *Linguistics and Philosophy* **3**:143–84.
- Nunberg, G. (1995), 'Transfers of meaning'. *Journal of Semantics* **12**:109–32.
- Ostler, N & B. Atkins (1992), 'Predictable meaning shift: some linguistic properties of lexical implication rules'. *Lexical Semantics and Knowledge Representation* **627**:87–100.
- Papafragou, A. (1996), 'On metonymy'. *Lingua* **99**:169–95.
- Pustejovsky, J. (1995), *The Generative Lexicon*. MIT Press. Cambridge, MA.
- Rabagliati, H., G. F. Marcus & L. Pykkänen. (2010), 'Shifting senses in lexical semantic development'. *Cognition* **117**:17–37.
- Rehder, B. (2003), 'Categorization as causal reasoning'. *Cognitive Science* **27**:709–48.
- Rehder, B. & R. Hastie. (2004), 'Category coherence and category-based property induction'. *Cognition* **91**:113–53.
- Rehder, B. & S. Kim. (2006), 'How causal knowledge affects classification: a generative theory of categorization'. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **32**:659–83.
- Rosch, E. & C. Mervis. (1975), 'Family resemblance: studies in the internal structure of categories'. *Cognitive Psychology* **7**:573–605.
- Schütze, H. (1997), 'Ambiguity Resolution in Language Learning'. CSLI Publications. Stanford, CA.
- Slooman, S., Love, B., & Ahn, W. (1998), 'Feature centrality and conceptual coherence'. *Cognitive Science* **22**:189–228.
- Strigin, A. (1998), 'Lexical rules as hypotheses generators'. *Journal of Semantics* **15**:163.
- Wilson, D. (2003), 'Relevance and lexical pragmatics'. *Italian Journal of Linguistics* **15**:273–92.
- Zipf, G. K. (1945), 'The meaning-frequency relationship of words'. *Journal of General Psychology* **33**:251–56.
- First submission: 28.06.2010
 Second submission: 06.09.2010
 Accepted: 03.01.2011