

Running Head: CONSTITUENT POLYSEMY IN COMPOUND INTERPRETATION

Constituent polysemy and interpretational diversity in attested English novel
compounds

Martin Schäfer¹ and Melanie J. Bell²

¹Eberhard Karls Universität Tübingen

²Anglia Ruskin University, Cambridge

Address for correspondence:

Martin Schäfer, Eberhard Karls Universität Tübingen, SFB 833, Nauklerstr. 35, 72074
Tübingen

Email: post@martinschaefer.info

Abstract

We explore variation in the interpretation of attested novel compound nouns in English, especially the contribution of constituent polysemy to this diversity. Our results show that effects of polysemy are pervasive in compound interpretation, contributing both to interpretational diversity and to perceived difficulty of interpretation. The higher the uncertainty about the concept represented by the head noun, based on existing compounds with that head, the greater the diversity of interpretations across speakers and the more difficult, on average, they find it to come up with a meaning.

Keywords: novel compound, compound noun, compound interpretation, constituent polysemy, constituent family

Because compounds are semantically underspecified, a previously-unseen compound presented in isolation has more than one possible interpretation (e.g. Ryder, 1994; Wisniewski, 1996). Compounds also vary in how difficult it is for people to come up with an interpretation and there is some evidence that compounds perceived as more difficult to interpret generate more variation in the interpretations produced (Coolen, Van Jaarsveld, & Schreuder, 1991). This variation is partly attributable to the unexpressed semantic relation between the constituent words. For example, nothing in the form *cinnamon stick* tells us that the intended meaning is ‘a dried strip of cinnamon bark’, broadly a stick OF cinnamon, rather than e.g. a stick FOR supporting the growth of a cinnamon tree (cf. *walking stick*) or a stick FOR holding small pieces of cinnamon, perhaps for purposes of infusion (cf. *cocktail stick*). However, compound interpretation also involves attributing an appropriate sense to each constituent. For example, to interpret *rubber plant* one not only has to establish a relation (e.g. plant PRODUCES rubber, plant MADE OF rubber or plant PROCESSES rubber) but also the meaning of each noun: does *plant* signify a living organism, an artefact in the shape of a living organism, or a factory?

This paper investigates the role of constituent ambiguity in the interpretation of previously-unseen (henceforth ‘novel’) compounds. Specifically, we aim to address the following questions:

1. How much diversity is there in the interpretation of attested novel compounds?
2. What role does constituent polysemy play in this diversity and in perceived difficulty of interpretation?

Two terminological notes are in order. Firstly, regarding the term ‘compound’: following Bell (2011), we classify all constructions of two nouns in which the first noun modifies the second as compounds, irrespective of orthography. We therefore refer to the items in our dataset as compounds, even though they are written with spaces. Secondly, regarding the term ‘polysemy’: although a distinction is sometimes made between ‘polysemy’ for the existence of multiple related senses and ‘homonymy’ for the existence of unrelated senses, in practice it is hard to draw a clear line between these two phenomena and we therefore use the term ‘polysemy’ for all cases of multiple word senses.

We first review the literature on compound interpretation and processing with particular reference to constituent polysemy. We then describe how we created a database of novel compounds, before presenting two studies addressing each of our questions in turn. The paper concludes with a discussion of our results.

Background

Constituent polysemy has received scant attention in the literature on compound interpretation, with most research focussing instead on the semantic relation between

constituents. Where studies have investigated the relationship between a compound's semantic relation and the semantics of its constituents, they have tended to address the issue of polysemy by attempting to circumvent it. Maguire, Wisniewski, and Storms (2010) found that the most accurate prediction of a compound's semantic relation is obtained by considering the semantic categories of both constituents. For example, if the first noun (N1, the modifier) represents a substance, and the second noun (N2, the head) represents an artefact, as in *steel pipe*, the most probable relation is MADE OF. However, the authors recognised that a given noun might belong to more than one category: e.g. *dog* has different senses belonging to the categories 'animal', 'food' (as in *hotdog*) and 'artefact' (as in *fire dog*). Their solution was to exclude constituents if fewer than 90% of their corpus occurrences fell into a single class. Ryder (1994) also tried to avoid ambiguous nouns in her compounds, but noticed that her participants nevertheless sometimes used different meanings of a constituent in their interpretations. In contrast, Wisniewski (1996) did not consider polysemy in the construction of his material and was therefore obliged to address it post-hoc by including the category 'noun construal' in his coding of participants' interpretations: this term was applied when a constituent was given a different meaning from what the author deemed to be 'the object typically named by the noun'. Such constricts usually involve a sense of N1 that designates a property of N2, as in *squirrel car* meaning 'a small fast car', or a sense of either constituent that designates a representation of something else, as in *car box* meaning 'a box that holds toy cars' or *stone squirrel* meaning 'a figure of a squirrel made in stone'.

The influence of constituent polysemy on compound semantics has been addressed tangentially in the literature on semantic transparency. According to Bell and Schäfer (2013), for example, each constituent word of a complex nominal (a compound or adjective-noun phrase) can have either a 'literal' or a 'shifted' sense, and interpretation of the whole construction depends upon which sense is selected. Similarly, Libben (2014) suggested that opaque compounds can arise through a process of 'morphological transcendence', when constituent meanings drift from the meaning of the original word, so that e.g. *key-* in *keyboard* has a different representation from either *-key* in *room key* or *key* as a free form. However, he only considered the issue of polysemy between morphological positions, and not within the same position, such as the difference in meaning between *key-* in *keyboard* and *key-* in *keyring*.

In the literature on lexical processing, we are only aware of two studies that have focussed explicitly on the role of constituent polysemy in complex nominals. Coolen, Van Jaarsveld, and Schreuder (1993) constructed novel Dutch compounds whose heads were ambiguous nouns with a dominant and a non-dominant (i.e. less frequent) meaning in the language generally. Each compound was judged to have an 'obvious' interpretation involving either the dominant or non-dominant sense of the head. In a priming study with lexical decision, using the compounds as primes, they found facilitatory effects only for the sense of the head used in the compound, whether or not

this was the dominant sense generally. They concluded that lexical access can be restricted to particular senses when the context is sufficiently constraining, and that each constituent of a noun-noun compound can constrain the possible interpretations of the other constituent. Mullaly, Gagné, Spalding, and Marchak (2010) investigated the interpretation of adjective-noun constructions with ambiguous adjectives, again using dominant and non-dominant senses of the ambiguous words. In this case, some facilitatory priming was found even across different senses of the adjectives. The authors concluded that lexical representation must involve some kind of core meaning that includes aspects of all more specialised senses, or is connected to them in some other way. Both the differing results of these two studies, and the fact that each one considered only two possible senses of a single constituent, underline the fact that there is still much to be discovered about the contribution of constituent polysemy to compound interpretation.

At a theoretical level, there has been little discussion of the role of lexical ambiguity in compound interpretation. Prominent theories have their origins in the study of conceptual combination, where concepts are taken as the starting point and the processes whereby phonological or orthographic forms are mapped onto concepts are therefore not considered. For example, the Relational Interpretation Competitive Evaluation theory (RICE, Spalding, Gagné, Mullaly, & Ji, 2010) suggests that compound interpretation proceeds in three possibly overlapping stages: firstly, the concept represented by N1 suggests various likely semantic relations; secondly, these possibilities are evaluated for compatibility with the concept represented by N2 to identify an appropriate relation at gist level; finally, this relational structure is elaborated using knowledge about the concepts involved. The theory assumes that identification of the concepts represented by the constituent nouns precedes identification of the relation between them, but does not spell out how the appropriate concepts are selected in cases where the constituent nouns are polysemous.

Creating a database

Materials

In previous research on novel compounds, researchers have usually created sets of experimental items to suit the particular objectives of the study in question. A representative example is Coolen et al. (1991), who constructed a set of compounds by combining simple nouns in accordance with their intuitions about the interpretability of the various combinations. Novel compounds were contrasted with lexicalised compounds, and lexicalisation was operationalised as being listed in a dictionary. Our approach is different in two important ways. Firstly, rather than inventing items ourselves, we used only attested compounds. This was because we do not assume that any given combination of two nouns can form an interpretable compound. Using

attested compounds ensured that our examples had occurred in natural language use at least once and were therefore in principle interpretable. Secondly, because we were interested in the diversity of possible interpretations, we wanted to avoid participants simply retrieving a meaning from memory: this could lead to people converging in their interpretations of a compound simply because they had all learnt the same meaning. However, lack of a dictionary entry does not reliably indicate that a compound will be unfamiliar. For example, neither *cereal box* nor *phone mast* nor *shoe polish* occur in the Oxford English Dictionary, but all are probably well known to very many speakers of the language. To address this issue, we used corpus frequencies: all items had to be attested in actual language use, but they needed to be so rare that the probability of participants having encountered them before was as small as possible. To this end, we used compounds that occurred exactly once in a very large corpus, namely ukWaC (Baroni, Bernardini, Ferraresi, & Zanchetta 2009), a corpus of more than 2 billion words from the .uk internet domain.

To identify candidate noun-noun strings, we generated all possible binary combinations of the compound constituents for which we had previously published semantic annotations (Bell & Schäfer 2016). We then selected only those combinations with a total ukWaC lemma frequency of one, summed across all possible spellings (British, American, spaced, hyphenated and concatenated). This yielded a set of 281 combinations, for which we extracted the corpus contexts. Combinations were then excluded if the two nouns did not form a compound in the context, or were part of a larger compound, or were marked as special by the use of quotes, or appeared to be part of a non-standard dialect. Candidates whose status was unclear were also excluded, as were any items in which N1 was an *-ing* form, N2 was a deverbal formation (e.g. *teacher*) or either constituent was itself a compound (e.g. *cocktail*). This led to a set of 60 novel compounds, in which many constituent nouns occurred several times. After further trimming the data so that no constituent occurred more than twice in total across head and modifier positions, we were left with a set of 45 items.

Participants

Because we were primarily interested in properties of the compounds, we took steps to reduce the amount of interpretational diversity that might result from sociolinguistic diversity amongst our participants. To this end, we recruited high school students aged 16-19 from two schools with the same catchment area. Although school pupils will have varied backgrounds, we assume that once they leave school their language experience will diversify even more widely, according to the life-choices they make. Using school students was thus an attempt to maximise homogeneity of linguistic experience. In addition, participants had to be monolingual native speakers of English who had grown up in the East of England and had no known language or hearing-related disability. Personal information, including gender, was collected using free-text fields. Half the

participants self-identified as ‘male’, and half as ‘female’. In total, 40 students participated in return for payment of £10.

Procedure

The study was designed as a computer-based questionnaire, presented using a Python 3 script (<https://www.python.org/>). For each compound, participants were first asked: ‘What does [COMPOUND] mean?’ After typing their interpretation into a free text box, they saw a second screen with the question: ‘How difficult was it for you to think of a meaning for [COMPOUND]?’ For this second question, an explicitly labelled Likert-type scale was used, with the labels: ‘Extremely difficult’, ‘Moderately difficult’, ‘Slightly difficult’, ‘Slightly easy’, ‘Moderately easy’ and ‘Extremely easy’. The study was self-paced because a pilot study had shown that when participants felt pressured for time, they tended to write ‘don’t know’ instead of coming up with an interpretation. Also the time required to put a meaning into words and type those words varied considerably between participants. We wanted to access participants’ interpretations equally, irrespective of their typing ability or linguistic fluency. Finally, we were particularly interested in variation between compounds and had no expectation that every item would require the same short time to interpret. Each participant saw a different random selection of 22 or 23 compounds, presented in a randomised order. The lists were created in pairs so that each pair of lists included every compound and no participant saw a constituent noun more than once. All data collection took place between January and April 2018 in the same purpose-designed psychology testing suite, in the presence of the first author of this paper. The full dataset is available at <https://XXXXX>.

Study 1: How much diversity is there in the interpretation of attested novel compounds?

Qualitative Coding of Free Paraphrases

Because we had no preconceptions about the interpretations that participants would come up with, we employed a bottom-up categorisation process to group the paraphrases into a set of distinct interpretations for each compound. Two coders, the second author of this paper and an experienced lexicographer, both native speakers of English, categorised the data. For each compound, the coders saw the 20 paraphrases in a different random order, and independently identified those paraphrases that they thought converged on the same interpretation. This was done using the constant comparative method; i.e. there was no a priori set of categories, but instead the coders introduced categories as needed to adequately capture the data. After completing the categorisation individually, the two coders resolved any discrepancies through discussion and consensus over a series of meetings. They also added a key in the form

of minimal definitions of the interpretations. For example, for the novel compound *car tower* the coders identified eight distinct interpretations. These are shown in (1), together with the number of participants agreeing on each interpretation.

- (1) a. vertical arrangement of multiple cars [10]
- b. car parking structure [4]
- c. tall building made of cars [1]
- d. horizontal arrangement of multiple cars [1]
- e. tower that looks out for cars [1]
- f. car transporter [1]
- g. vehicle for towing cars [1]
- h. tow bar [1]

The coding keys (1c) to (1h) all correspond to interpretations given by a single participant. In contrast, (1a) and (1b) capture ten and four paraphrases, respectively. To illustrate this grouping procedure, the actual paraphrases coded with coding key (1b) are given in (2).

- (2) a. *A tower for car parking*
- b. *Car park*
- c. *A tall parking building*
- d. *A multi-story carpark*

If a paraphrase clearly expressed two different meanings (e.g. the participant had written ‘or’ in the free text), the paraphrase was split into the two senses, leading to 24 additional paraphrases overall. If a paraphrase was ambiguous, it was coded as unclear between the two possible interpretations and subsequently excluded from analysis. In addition, one paraphrase of *ivory wall* was excluded because the participant had misread the first noun as *ivy*. This left us with a total of 916 valid paraphrases.

For all but 10 out of the 45 compounds, the coders identified cases where certain interpretations were subordinates of other more general interpretations within the data. Thus, for these compounds, there is both a fine-grained coding, and a coding at the superordinate level. In some cases, the superordinate coding is only minimally different from the fine-grained coding. For example, for *cloud floor*, the paraphrase (3a) was coded as a superordinate to (3b).

- (3) a. the surface of a cloud
- b. the bottom of a cloud

For other compounds, the difference between fine-grained and superordinate coding is greater. For example, Table 1 shows the coding for *gold student* and the proportion of paraphrases falling within each interpretation. For this paper, we were mainly interested in general patterns in compound interpretation, and therefore used the superordinate coding in our analyses.

Table 1: *Fine-grained and Superordinate Coding for Gold Student*

Gloss	Fine-grained Coding		Superordinate Coding	
	Code	Proportion	Code	Proportion
very good student	A	0.18	A	0.91
high-achieving student	A1	0.32		
best student	A2	0.18		
hard-working student	A3	0.14		
uniquely-gifted student	A4	0.05		
well-behaved student	A5	0.05		
teacher's pet	B	0.09	B	0.09

Quantitative Measures of Interpretational Diversity

For each of the 45 compounds, we calculated the following four measures:

1. Convergence: Proportion of participants who gave a non-unique interpretation, i.e. whose paraphrase corresponded to the same interpretation as at least one other participant.
2. Maximum agreement: Proportion of participants whose paraphrases corresponded to the most frequent interpretation.
3. Spread: Number of distinct interpretations divided by the number of valid paraphrases.
4. Interpretational entropy (H), calculated using the formula in (4), where p_i is the proportion of paraphrases belonging to a given interpretation:

$$(4) H = -\sum_{i=1}^n p_i \log p_i$$

This is a measure of the overall level of uncertainty about the interpretation of a compound. It is highest when there are many equally frequent interpretations, i.e. when it is difficult to predict how the compound will be interpreted, and lower when there are few possible interpretations, or one interpretation is more frequent than others.

To illustrate these measures, consider the coding of *car tower* introduced in (1). Of the eight interpretations, six are unique, with the two non-unique interpretations shared by 10 and four participants, respectively. This yields the values in (5).

- (5) a. convergence: $14/20 = .70$
 b. maximum agreement: $10/20 = .50$
 c. spread: $8/20 = .40$
 d. interpretational entropy: 2.26

Results

Figure 1 shows the distributions of the four diversity measures, and Table 2 gives their pair-wise correlation coefficients. The four measures are highly correlated with one another and all are continuous across the 45 compounds, with only maximum agreement clearly deviating from a normal distribution. This suggests that studies that operationalise interpretational diversity as a binary contrast, though they do so for clear methodological reasons, may be missing a large portion of the possible variation. Particularly in the graph for maximum agreement (Figure 1 top right panel), it can be seen that interpretational diversity was high. Even using the superordinate coding, which effectively reduces the diversity values, only 11 compounds received majority interpretations in the sense that more than 50% of paraphrases belonged to the most frequent interpretation. For another 11 compounds, the most frequent interpretation was shared by at most 25% of participants. This suggests that novel compounds with a clearly dominant context-free interpretation are probably atypical. On the other hand, there was no compound for which every participant gave a different interpretation.

Table 2: *Pair-wise Pearson Product-moment Correlation Coefficients for the four Diversity Measures based on the Superordinate Coding ($p < .01$ for all pairings)*

	Maximum Agreement	Spread	Interpretational Entropy
Convergence	0.47	-0.93	-0.75
Max. Agreement		-0.71	-0.91
Spread			0.92

Table 3 shows the top and bottom five compounds ranked by interpretational entropy. For the top three, *search tower*, *memory tank* and *gold student*, 81%, 85% and 91% of the participants respectively agreed on an interpretation, leaving a clear gap to the next item, *call test*, at 65%. These top three items are the clear outliers in the distributions of maximum agreement and interpretational entropy in Figure 1. At the superordinate

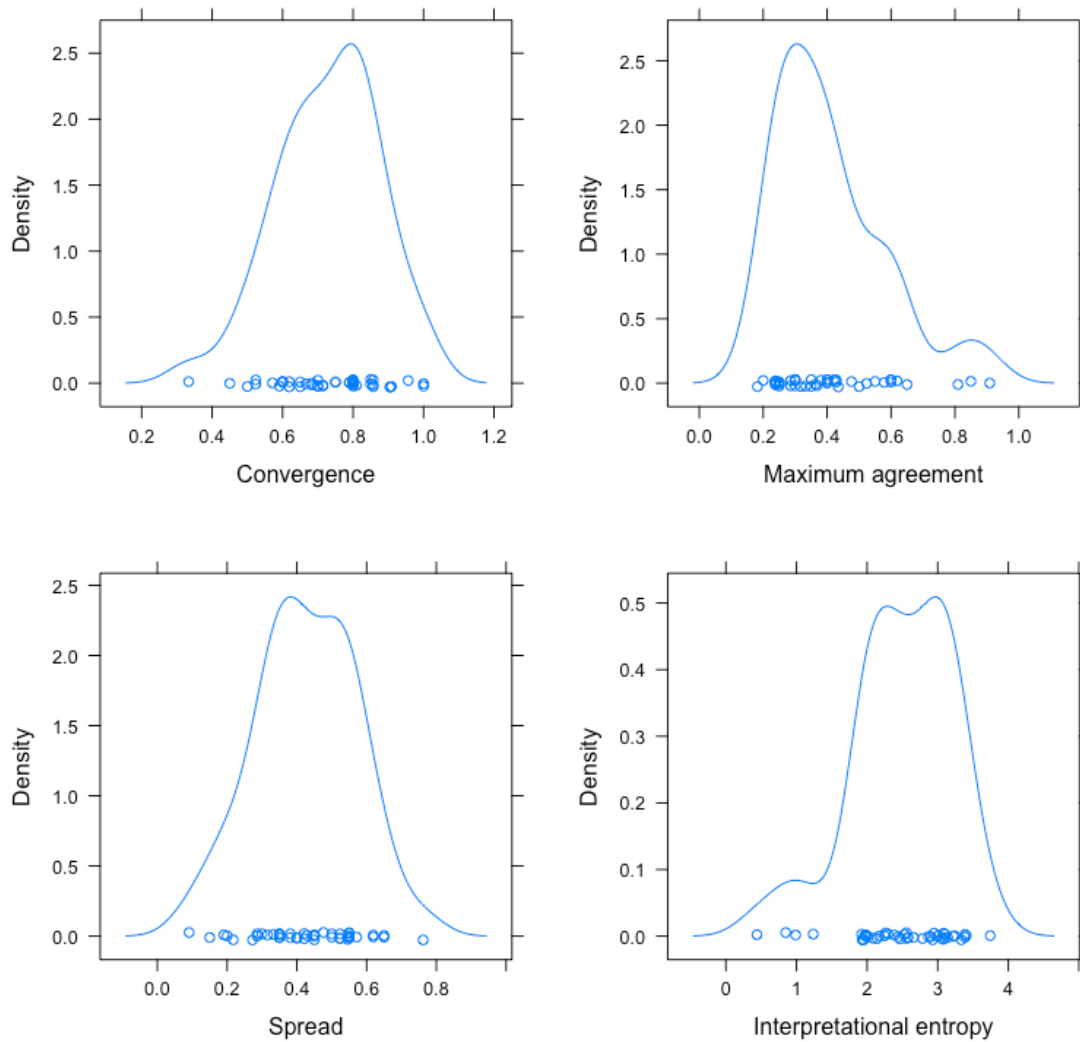


Figure 1. Density plots for the interpretational variety measures based on the superordinate coding: convergence is the proportion of participants who gave a non-unique interpretation, maximum agreement is the proportion of participants who gave the most frequent interpretation, spread is the number of distinct interpretations divided by the number of valid paraphrases, and interpretational entropy captures the distribution of the probabilities of the interpretations.

level, neither *gold student* nor *call test* had any interpretation that was not given by at least two participants, hence their score of 1.0 for convergence. *Gold student* also showed the lowest interpretational entropy and the lowest spread. *Face room* had the highest spread and highest entropy but lowest convergence. Table 4 shows the three significant correlations between the mean difficulty rating of each compound and the interpretational diversity measures. Average perceived difficulty correlates positively with spread and interpretational entropy, and negatively with convergence. In other

words, the more difficult, on average, participants found it to interpret a compound, the greater the diversity in the interpretations they produced.

Table 3: *Diversity Measures for the Top and Bottom Five Compounds Ranked by Interpretational Entropy*

Compound	Convergence	Max. Agreement	Spread	Int. Entropy
<i>gold student</i>	1.00	0.91	0.09	0.44
<i>memory tank</i>	0.85	0.85	0.20	0.85
<i>search tower</i>	0.90	0.81	0.19	0.99
<i>call test</i>	1.00	0.65	0.15	1.24
<i>snake tank</i>	0.80	0.40	0.30	1.92
<i>diamond pool</i>	0.68	0.18	0.55	3.33
<i>monkey ring</i>	0.52	0.24	0.62	3.39
<i>radio shift</i>	0.52	0.24	0.62	3.39
<i>ground rate</i>	0.50	0.25	0.65	3.40
<i>face room</i>	0.33	0.24	0.76	3.74

Table 4: *Correlations between Mean Difficulty Rating and Interpretational Diversity Measures (based on the Superordinate Coding)*

Measure	Correlation with Mean Difficulty	
	Pearson's r	p -value
Spread	0.47	< 0.01
Convergence	-0.47	< 0.01
Internal Entropy	0.41	< 0.01

Study 2: What role does constituent polysemy play in the diversity of compound interpretation?

Quantitative analysis

In order to investigate the relationship between interpretational diversity and constituent ambiguity, we needed to operationalise polysemy. For each compound constituent in our data, we therefore calculated the following four measures:

1. Number of senses in the language generally, including all parts of speech
2. Number of noun senses in the language generally
3. Number of senses in the positional constituent family (the set of compound types with the given constituent in the same position)
4. Entropy of senses in the positional constituent family, using the formula in (4).

The first two of these measures were extracted from the lexical database WordNet (<https://wordnet.princeton.edu/>). The third and fourth measures came from the annotated set of compound families on which the selection of our data was based (Bell & Schäfer 2016). In that database, the compound types in the positional constituent families of various nouns are coded for the word senses of their constituents.

For each of the four polysemy measures per constituent, we calculated the pair-wise Pearson product-moment correlation coefficient with each of our four interpretational diversity measures. The only polysemy measure that consistently reached significance was the entropy of senses in the positional constituent family of N2. This variable showed moderate positive correlation with both Spread ($r(43) = .33, p = .028$) and Interpretational entropy ($r(43) = .31, p = .038$), and moderate negative correlation with Convergence ($r(43) = -.31, p = .037$). In other words, the greater the uncertainty about the probable sense of a noun when it is used as a compound head, the greater the diversity in the interpretations produced for a novel compound with that noun as its head. The entropy of senses in the positional constituent family of N2 was also moderately positively correlated with mean difficulty rating ($r(43) = .42, p = .005$).

Qualitative analysis

To understand the role of constituent polysemy in more depth, we examined a sample of eight novel compounds, in pairs that had similar interpretational entropy in the superordinate coding and represented all four quartiles of the distribution. The first pair are *memory tank* and *search tower*, with entropy values of 0.8 and 1.0 respectively. For each of these compounds, the coders distinguished only four superordinate meanings, shown in (6) and (7) respectively.

- (6) *memory tank*
- a. place where memories are stored [17]
 - b. memories [1]
 - c. people used as memory repositories [1]
 - d. database [1]

For *memory tank*, interpretation (6a) was completely dominant, accounting for 85% of the paraphrases. It might be regarded as a gist interpretation, and in the fine-grained coding it was split into three subcategories: the brain, part of the brain and an inanimate container of memories. Each of the minority interpretations (6b-d) involves a slightly different interpretation of one or the other constituent. In (6d), *tank* was taken to represent a set of inanimate, rather than mental, data. In (6c), the sense of *tank* is metaphorical, perhaps in analogy to *think tank*. In (6b), at first sight, the contribution of *tank* seems to have disappeared completely, though it is possible that the participant was making an analogy with *data bank*.

- (7) *search tower*
- a. tower for searching from [17]
 - b. lighthouse [2]
 - c. filing system [1]
 - d. game [1: 'A large puzzle game tower (puzzle solving game)']

For *search tower*, there was again a dominant interpretation (7a, 81%) and the different readings involved polysemy of the head. In both (7a) and (7b), *tower* has its 'building' sense, whereas in (7d) it appears to have the sense of 'something large', perhaps taller than it is wide. The sense in (7c) is not completely clear but may be 'stack', e.g. of drawers or filing cabinets.

The role of polysemy becomes even more apparent when we move to the second quartile of interpretational entropy, exemplified by *car tower* (1) and *couch room* (8). Although the two dominant interpretations of *car tower*, 'vertical arrangement of multiple cars' and 'car parking structure' are clearly distinguished by a different relation, i.e. MADE OF vs. FOR, they also involve, respectively, the 'building' and 'pile' senses of *tower*. Constituent polysemy comes yet further to the fore in the final two interpretations, 'truck used to tow cars' and 'tow bar'. In these cases, the participants analysed *tower* as an instrumental deverbal *-er* nominal, rhyming with *lower*.

- (8) *couch room*
- a. living room [7]
 - b. room with a couch or sofa in it [5]
 - c. room with couches [5]
 - d. place to relax [3]
 - e. waiting area [1]
 - f. amount of space on a sofa [1]

For *couch room*, the coders distinguished six interpretations, with all except one seeming to build on the same constituent senses. Only (8f) clearly uses a different sense of *room*, i.e. ‘unoccupied space’ rather than ‘part of a building’. Thus, for both *car tower* and *couch room* we found a clear effect of constituent polysemy, but restricted to N2 and only resulting in minority interpretations.

The third quartile is represented by *snail model* and *rat order*. Both showed polysemy in N1 as well as N2, with some interpretations exploiting metaphorical meanings of the modifiers. For *snail model*, the different modifier senses are illustrated in (9). Besides the animal reading of *snail*, the term was taken to indicate speed, size or shape.

- (9) a. slow model [one that takes a long time to build] [2]
- b. slowed-down presentation of something [3]
- c. very small model [1]
- d. specific way of designing something [starting in the centre and expanding outwards in spirals] [1]

Model occurred with at least three different senses, most frequently variations on the ‘representation’ meaning, but also with the senses ‘person employed to wear clothes for display’ (10a) and ‘exemplar of some excellence’ (10b).

- (10) a. snail that’s fashion model [1]
- b. attractive snail [1]

For *rat order*, participants used the animal sense of the modifier as well as the metaphorical sense of ‘contemptible person’, and in fact anything unpleasant (11).

- (11) a. hierarchy based upon unpleasantness [1]
- b. unpleasant order [1]

Order showed even greater polysemy, with at least four different meanings being used. The dominant senses were ‘logical arrangement’ (12a) and ‘social hierarchy’ (12b), followed by ‘organised group’ (12c) and ‘command’ (12d).

- (12) a. sequence into which rats are arranged [6]
- b. natural hierarchy [4]
- c. organised group of rats (e.g. an army) [2]
- d. harsh command (vocal quality) [1]

In the two examples from the fourth quartile, *diamond pool* and *monkey ring*, the role of constituent polysemy increased still further. *Diamond pool* had twelve superordinate interpretations, involving polysemy of both N1 and N2. *Diamond* was taken to represent either the precious stone (e.g. ‘a pool filled with diamonds’) or the geometric figure (e.g. ‘a diamond-shaped pool’), and the precious-stone meaning was used as the basis for several shifts, involving either the concepts of brilliance and clearness or of wealth and extravagance (13).

- (13) a. glistening pool [2]
- b. pool of clear water [3]
- c. expensive pool [1]
- d. very big swimming pool [1]

For *pool*, the most frequent reading was ‘body of water’, including the more specific variants ‘rock pool’ and ‘swimming pool’. But participants also employed the senses ‘supply of resources’ (14a) and ‘group of people’ (14b).

- (14) a. large supply of diamonds [1]
- b. wealth/wealthy people [1]

Monkey ring, with thirteen interpretations overall, likewise showed variation in the readings of both N1 and N2. Most interpretations used the animal sense of *monkey*, but participants also exploited the metaphorical sense ‘mischievous person’ (15a) and the verbal sense ‘to mess about’ (15b).

- (15) a. people behaving badly [3]
- b. being silly with friends [1]

Ring was predominantly used with either the shape or jewellery senses (16a and 16b), but the ‘group of people’ sense and ‘performance space’ senses also occurred (15 and 16c).

- (16) a. monkeys in a circle configuration [5]
- b. ring with monkey on it [3]
- c. circus ring [1]

Across our set of 45 novel compounds, there were only two where we could be reasonably certain that only one meaning of the head was used: *gold student* and *health song*. In both these cases, different senses of the modifier were used. So overall, constituent polysemy was evident to a greater or lesser extent in the range of interpretations produced for every compound in the dataset. Some compounds showed clear association between specific constituent senses and the semantic relation between constituents. For example, when *room* in *couch room* was taken to mean ‘part of a building’, the relation could be classified as N1 IN N2. But when *room* was interpreted as ‘space’, as in ‘the amount of space on a sofa’ (8f), the relation was reversed, and the room was seen to be located on the couch. The senses of N1 and N2 also interacted, as can be nicely seen in some of the compound pairs with a shared constituent. For example, while interpretations of *gold cow* involved the metal sense of *gold*, amongst other possibilities, *gold* in *gold student* excluded the metal sense. Similarly, *song* in *health song* was consistently interpreted as ‘short musical composition’, whereas in *kangaroo song* it was also interpreted as ‘form of animal communication’.

Discussion

Our results show that the inherent ambiguity in novel English compound nouns leads to considerable diversity in their context-free interpretations. Nevertheless, some produce more interpretational diversity than others and, for any given compound, some interpretations are more frequent than others. None of the novel compounds in our dataset were interpreted differently by every participant: all had at least one interpretation that was shared by at least two people. On the other hand, for all but two of the compounds, at least one person gave an idiosyncratic interpretation not shared by anyone else.

The pattern of our results is reminiscent of the pattern found in norming studies of word association, where participants are asked to write the first word that comes to mind that is meaningfully related or strongly associated with a stimulus word. Nelson, McEvoy and Schreiber (1998) report that, when a given set of words is re-normed with different participants, responses that were originally given by at least two people are found to be ‘reliable’ in the sense that they usually recur across groups of speakers. However, responses given by just one person are found to be ‘unreliable’ in that, although they make sense, they tend to be different every time. Comparing the data reported in this paper with the data from our pilot study, we also find relatively common recurring interpretations plus a long tail of idiosyncratic responses unique to individual speakers. In reaching an interpretation of a previously-unseen compound, a speaker can draw on many kinds of information including, but not limited to, their experience of how the constituent words are used elsewhere in the language, their experience of other compounds, their experience of the world and their encyclopaedic knowledge. It is possible that frequent interpretations draw on information at a more general level using more widely-shared knowledge, including more frequent linguistic patterns, whereas idiosyncratic interpretations draw on more specific knowledge and experience, or on more recent events. For example, in the context of our data collection, a participant’s interpretation of any compound except the first could be influenced by their interpretation of the compounds they saw earlier in experiment.

We considered the possibility that our participants might produce idiosyncratic responses if they misunderstood the experiment as a test of their creativity. To investigate this, we calculated the number of times each participant gave a unique interpretation, i.e. one not given by any other participant. Although some participants gave more unique interpretations than others, no-one gave exclusively unique interpretations, while every participant gave a unique interpretation for at least one item. A Shapiro-Wilk test showed that the proportion of unique interpretations per participant did not depart significantly from a normal distribution ($W = .96, p = .20$). Overall, we have no reason to believe that any of the participants had misunderstood the task or were producing deliberately unusual or creative responses. In the context of the norming studies mentioned above, Nelson et al. (1998) found that the number of different words produced by a cue, regardless of whether they are given by two or more participants or

by a single participant, is a highly reliable statistic, i.e. consistently replicable across studies, even though the specific idiosyncratic responses vary. If our data follow a similar pattern, as we expect they do, then it seems likely that our diversity measures reliably reflect some underlying property or properties of the compounds concerned.

Study 2 showed that constituent polysemy is one compound property that contributes to interpretational diversity and plays a pervasive role in the interpretation of novel English compounds. For all 45 compounds in our dataset, participants produced interpretations involving more than one sense of at least one constituent. In the quantitative analysis, polysemy of the head was a significant predictor both of interpretational diversity and of perceived difficulty of interpretation. The higher the uncertainty about the concept represented by the head noun, based on existing compounds with that head, the greater the diversity of interpretations across speakers and the more difficult, on average, they found it to come up with a meaning.

Nevertheless, despite the clear relationship between compound interpretational diversity and polysemy of the constituents, we found only moderate levels of correlation between the relevant measures. Clearly, factors other than constituent polysemy also contribute to the overall ambiguity of a novel compound. One obvious candidate is the semantic relation between constituents, which might conceivably vary even for given senses of the modifier and head. We also found only moderate correlation between perceived difficulty of interpretation and either interpretational diversity or polysemy of the head. Again, this indicates that, although greater ambiguity makes a compound more difficult to interpret, other factors also come into play. Such factors might conceivably include the availability or otherwise of a relevant analogical pattern in the language, or the availability of an appropriate contextual schema in the experience of the participant. In other words, even if a compound had only one possible reading, and was therefore not ambiguous, that reading might be more or less difficult to reach.

To what extent do our findings support Libben's (2014) notion of morphological transcendence? The fact that, amongst our polysemy measures, entropy of senses in the N2 positional family is the best predictor of interpretational diversity, suggests that position-specific senses may indeed be the most relevant for compound interpretation. However, it is unclear to what extent these senses also occur as free forms. Are the senses of N2 in the positional family a subset of its senses as a free form, or a different set? In our data, the number of senses of N2 in the positional constituent family is strongly positively correlated with the number of senses in the language generally ($r(43) = 0.75, p < .001$) and especially the number of senses classed as nouns ($r(43) = 0.84, p < .001$). If compound-specific senses had drifted to the extent of being independent of free-form senses, then we would not necessarily find such a strong correlation. In order to resolve this issue, a more detailed study of the distribution of alternative senses would be required, perhaps using a sense-disambiguated corpus.

How do our results fit with the RICE theory of conceptual combination as applied to the interpretation of compound nouns? Gagné and Spalding (2014) suggest that, at least for modifiers, selection of specific senses might take place at the elaboration stage, i.e.

after a gist interpretation has been established. But, according to the theory, in order to reach a gist interpretation, the concept represented by the modifier has to suggest semantic relations to be evaluated for compatibility with the concept represented by the head. In other words, if the theory is correct, then some mapping of the modifier to a concept has to take place from the earliest stage of interpretation, closely followed by the head noun. One possibility is that each noun initially activates some sort of core conceptual representation, along the lines suggested by Mullaly et al. (2010), which is narrowed down once a gist interpretation has been established. For some items in our data, it seems plausible that several or perhaps all senses might have a shared core. One such example is *diamond*, where interpretations other than the ‘gem’ sense are clearly related to perceived characteristics of the gem. However, for other items a shared-core sense seems much less plausible. For example, interpretations of *rush trip* involved both the ‘hurry’ sense of *rush* (‘a hurriedly-planned trip’) and the ‘vegetation’ sense (‘a trip to the rushes (in a nature reserve)’). In this case, it is quite hard to formulate what these two senses might have in common, beyond the shared form.

Rather than assuming that nouns activate shared-core concepts related to all their possible senses, an alternative formulation of RICE theory is that the modifier activates a variety of concepts, each of which suggest relations in parallel. These would then be evaluated by all concepts associated with the head. In many cases this would lead to a large number of possibilities; however, two factors could constrain this potential proliferation. Firstly, in actual language use, the context would make certain constituent senses much more likely than others, so that not all possible concepts would be activated to the same extent. Secondly, as suggested by Gagné and Spalding (2014) for semantic relations, weak competitors might require very little negative evidence, whether from the other constituent or the wider context, to rule them out. At first sight, this theoretical proposal might seem to be inconsistent with the results of Mullaly et al. (2010), who found facilitatory priming across different senses of adjectives in adjective-noun phrases. However, since that study only looked at adjectives, and especially since Coolen et al. (1993) did not find priming across senses in Dutch noun-noun compounds, there remains the possibility that the property concepts typically signified by English adjectives are differently represented in the mental lexicon than the entity concepts typically signified by nouns.

In summary, we have shown that effects of constituent polysemy are pervasive in the interpretation of novel compounds, contributing both to interpretational diversity and to perceived difficulty of interpretation. This polysemy has not so far been thoroughly investigated or incorporated into any theory of compound interpretation. However, as Mullaly et al. (2010) point out, “developing a comprehensive theory of compounding will ultimately involve understanding how the required sense of a constituent is identified”.

Acknowledgements

We gratefully acknowledge funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No. 752436. Many thanks to David Atter from Hills Road Sixth Form College for his support of our project, and to Diane Nicholls, the second coder of our data. The paper has profited greatly from the comments of two anonymous reviewers, the advice of the editors of this issue, and feedback from the audience at the Workshop on Psycholinguistic Aspects of Complex Words, 52nd Annual Meeting of the Societas Linguistica Europaea, Leipzig 2019.

References

- Baroni, M., Bernardini, S., Ferraresi, A. & Zanchetta, E. (2009). The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3), 209–226.
- Bell, M. J. (2011). At the boundary of morphology and syntax. Noun noun constructions in English. In A. Galani, G. Hicks, & G. Tsoulas (Eds.) *Morphology and its Interfaces* (pp. 137–168). Amsterdam: Benjamins.
- Bell, M. J. & Schäfer, M. (2013). Semantic transparency: challenges for distributional semantics. In *Proceedings of IWCS 2013 Workshop Towards a Formal Distributional Semantics*, (pp. 1–10). Association for Computational Linguistics.
- Bell, M. J. & Schäfer, M. (2016). Modelling semantic transparency. *Morphology* 26(2), 157–199.
- Coolen, R., Van Jaarsveld, H. J. & Schreuder, R. (1991). The interpretation of isolated novel nominal compounds. *Memory & Cognition* 19(4), 341–352.
- Coolen, R., Van Jaarsveld, H. J. & Schreuder, R. (1993). Processing novel compounds: Evidence for interactive meaning activation of ambiguous nouns. *Memory & Cognition* 21(2), 235–246.
- Gagné, C. L., & Spalding, T. L. (2014). Conceptual composition: The role of relational competition in the comprehension of modifier-noun phrases and noun-noun compounds. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 59 pp. 97–130). Elsevier Inc.: Academic Press.
- Libben, G. (2014). The nature of compounds: A psychocentric perspective. *Cognitive Neuropsychology* 31(1-2), 8–25.
- Maguire, P., E. Wisniewski, & G. Storms (2010). A corpus study of semantic patterns in compounding. *Corpus Linguistics and Linguistic Theory* 6(1), 49–73.
- Mullaly, A., Gagné, C., Spalding, T. L. & Marchak, K. (2010). Examining ambiguous adjectives in adjective-noun phrases: Evidence for representation as a shared core-meaning with sense specialization. *The Mental Lexicon* 5. 87–114.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. <http://www.usf.edu/FreeAssociation/>.
- Ryder, M. E. (1994). *Ordered chaos: The interpretation of English noun-noun compounds*. University of California Press.
- Spalding, T. L., Gagné, C. L., Mullaly, A. & Ji, H. (2010). Relation-based interpretation of noun-noun phrases: A new theoretical approach. In S. Olsen (Ed.), *New Impulses in Word-Formation*, (pp. 283–315). Hamburg: Buske.
- Wisniewski, E. J. (1996). Construal and similarity in conceptual combination. *Journal of Memory and Language* 35(3), 434 – 453.