

## The regularity of polysemy patterns in the mind: Computational and experimental data

**Alizée Lombard**, Université de Fribourg, CH, [alizee.lombard@unifr.ch](mailto:alizee.lombard@unifr.ch)

**Anastasia Ulicheva**, Royal Holloway University of London, UK, [anastasia.ulicheva@gmail.com](mailto:anastasia.ulicheva@gmail.com)

**Maria Korochkina**, Royal Holloway University of London, UK, [Maria.Korochkina@rhul.ac.uk](mailto:Maria.Korochkina@rhul.ac.uk)

**Kathleen Rastle**, Royal Holloway University of London, UK, [Kathy.Rastle@rhul.ac.uk](mailto:Kathy.Rastle@rhul.ac.uk)

Linguists have often observed that the sense extensions in polysemous words follow patterns. Yet, these patterns have rarely been quantified, and it is unknown whether language users are sensitive to them. We developed four regularity metrics, focusing in this initial study on metaphor patterns that apply to nouns. We further tested adult English speakers' capacity to understand new senses in an acceptability judgement task. We compared novel senses that followed a metaphor pattern against novel senses that did not respect any pattern. Our results showed that novel senses were judged as more acceptable when they were part of a polysemy pattern as opposed to when they were not. We also assessed whether acceptability judgements were influenced by the degree of regularity of the pattern that they follow. The results confirmed the psychological validity of degree of regularity as a measure: the more regular the polysemy pattern, the more acceptable the new sense following that pattern. Regularity metrics that captured the *consistency* with which a pattern is instantiated were more successful in predicting acceptability ratings than regularity metrics that captured the *number of times* a pattern is instantiated. These results motivate future psycholinguistic studies investigating the influence of regularity on learning, processing, and storage of polysemes in a more nuanced way than has been possible previously.



## 1. Introduction

Lexical ambiguity is one of the most fascinating properties of language. The same form can have several meanings; these meanings may be unrelated *homonyms* (e.g., dog *bark* versus tree *bark*), or they may be related *polysemes* (e.g., *run* for office versus *run* a company). For decades, research has probed why languages tolerate ambiguity (e.g., Wittgenstein, 1958), and how the cognitive system resolves the challenges it poses for language comprehension (e.g., Johnson-Laird, 1987; Swinney, 1979). This article focuses on a relatively understudied aspect of polysemy. Specifically, we ask whether there are patterns that govern the process of sense extension in polysemy, how these can be quantified, and whether adult language users are sensitive to them.

The different senses of polysemous words are most commonly linked by two types of lexical construct. One such construct is *metaphor*, where different senses have an analogical relationship, because they refer to something that is similar in form, function, or behaviour (e.g., *pig* ‘animal’ and *pig* ‘dirty person’). Alternatively, senses can be linked through *metonymy*, a construct in which different senses have a logical relationship, such as action-agent, part-whole, container-content, or cause-effect (e.g., *tin* ‘metal’ and *tin* ‘object made of tin’). Importantly, metaphor and metonymy are both constructs that allow the creation of new words by extending the meanings of already existing words, thus bypassing the need to create new lexical forms (Gibson et al., 2019). Polysemy provides a highly efficient mechanism for lexical creativity, and it is therefore unsurprising that its use is widespread in natural languages (Srinivasan & Rabagliati, 2015). Moreover, since there is some evidence that meaningful relationships between senses facilitate the learning of new senses (Fang et al., 2017; Rodd et al., 2012), it may be that polysemy also provides a vehicle for efficient word learning.

Polysemy can also be characterised in terms of the regularity of the underlying construct. Consider, for instance, *pig*, *wolf*, and *shark*. These words all have a base meaning, ‘animal’ (SENSE 1), and another, less frequent meaning indicating a person: ‘dirty person’ for *pig*, ‘lonely person’ for *wolf*, and ‘aggressive person’ for *shark* (SENSE 2). Because many other words in English exhibit the same type of extension (*bird*, *chicken*, *leech*, *sheep*, etc.), the pattern ANIMAL → PERSON can be described as regular (see also Apresjan, 1974). On the other hand, if a semantic extension appears only in one polysemous word (e.g., ANIMAL → COMPANY as in *unicorn* ‘fantastic animal’/‘startup company valued at over 1 billion’), it can be considered irregular. Previous discussion of polysemy has sometimes gone further than a binary distinction between regular and irregular patterns (Barque et al., 2018; Copestake & Briscoe, 1995; Dölling, 2020; Nunberg, 1995; Nunberg & Zaenen, 1992; Pustejovsky, 1995); for example, the pattern PHYSICAL PROPERTY → PSYCHOLOGICAL PROPERTY (*softness*, *solidity*, *volatility*, ...) intuitively seems more regular than the pattern ANIMAL → PERSON (*pig*, *wolf*, *shark*, ...). However, to date, few attempts have been made to assess or quantify the graded nature of polysemy regularity.

There has been substantial interest in how the different types of lexical ambiguity influence language processing (for an overview, see Eddington & Tokowicz, 2015; Falkum & Vicente, 2015; and see also Duffy et al., 1988; Frazier & Rayner, 1990; Klepousniotou, 2002; Pylkkänen et al., 2006), but work investigating how the regularity of polysemy patterns influences processing is very limited. In terms of types of ambiguity, one of the critical findings in this literature is that the unrelated senses of ambiguous words (e.g., dog *bark* versus tree *bark*) must have separate representations in the mental lexicon, while the related senses (such as *run* for office versus *run* a company) must have overlapping mental representations (e.g., Brocher et al., 2018; Frisson, 2009; Frisson & Pickering, 1999; Rodd et al., 2002). Studies have also found a stronger relationship between senses related through metonymy (such as *tin* ‘material’ versus *tin* ‘object’) than between those related through metaphor (such as *pig* ‘animal’ versus *pig* ‘dirty person’; Klepousniotou & Baum, 2007; Klepousniotou et al., 2008; Klepousniotou et al., 2012; Lopukhina et al., 2018; Yurchenko et al., 2020). In terms of the regularity of polysemy patterns, some studies have investigated whether senses resulting from irregular extensions are processed differently from those resulting from regular extensions (Brocher et al., 2018; Rabagliati & Snedeker, 2013), while others have examined the extent to which polysemy patterns are shared across languages (Srinivasan & Rabagliati, 2015). However, these studies have tended to characterise different forms of polysemy in a categorical manner; for example, by associating regular polysemy with metonymy and irregular polysemy with metaphor (e.g., Apresjan, 1974; Brocher et al., 2018; Klepousniotou et al., 2012).

There have been various approaches to quantifying aspects of polysemy in a graded manner; however, most of these measures have focused on aspects of *single words* such as sense dominance (Gilhooly & Logie, 1980; Twilley et al., 1994) and sense uncertainty (e.g., Filipović Đurđević & Kostić, 2023) rather than properties of polysemy *patterns* such as regularity. To date, only one approach to quantifying polysemy regularity in a non-categorical manner has been developed. Srinivasan and Rabagliati (2015) asked 33 speakers of 14 languages to judge the extent to which 27 polysemy patterns found in English are also found in other languages. They found that patterns that had more similar senses across languages were also more generative, in the sense that novel senses that followed these patterns were judged as more acceptable. For example, in English, words like *chicken* and *lamb* can label both the animal and its meat (pattern ANIMAL → MEAT); this pattern is attested in many other languages and is thus highly generative. In contrast, while, in English, words such as *tin* and *glass* can denote both the material and the artefact made of this material (pattern MATERIAL → ARTEFACT), in many other languages examined in Srinivasan and Rabagliati (2015), the words denoting these materials did not denote the same artefacts (e.g., in French, *étain* denotes the same silvery-white metal the English word *tin* refers to, but, unlike *tin*, it cannot be used to refer to an airtight container made of tinfoil, and, in Russian, unlike English, the word that refers to the material ‘rubber’ can also refer to a car tyre), and in

other languages, this pattern was not present at all, suggesting that it is less generative. Critically, Srinivasan and Rabagliati’s approach to quantifying polysemy patterns was based on the notion of *cross-linguistic regularity*. One advantage of this approach is that it offers novel insights into why conceptual structure makes some sense relations easier to grasp than others, and hence, why particular polysemy patterns are attested across multiple languages. Yet their approach does not allow one to quantify the strength of specific polysemy patterns *within a language*. Moreover, because their metrics are based on the linguistic intuitions of a small number of informants for each language, the reproducibility of their results is difficult to attest. It is thus more desirable to quantify finer gradations of polysemy regularity using information that can be derived from corpora, yet, we are not aware of any research that has attempted to do that.

The research presented in this article therefore had two aims. Our first aim was to investigate how to quantify polysemy regularity in a graded manner and by means of corpus-derived metrics, focusing in this initial investigation on metaphor patterns that apply to nouns. We propose four potential metrics of regularity and describe the methods used to compute them. Our second aim was to test adults’ sensitivity to polysemy regularity using these four metrics. We tested adults’ capacity to understand new senses in an acceptability judgement task involving existing and novel polysemy extensions (e.g., ‘the knee of a mountain road’ derived from the BODY PART → OBJECT PART pattern). We then used our four metrics of regularity to assess whether the acceptability of these novel extensions was influenced by the degree of regularity of the polysemy patterns from which they were derived. In the following sections, we first describe the four measures of regularity that we developed and then report an experimental study in which we assessed speakers’ sensitivity to pattern regularity.

## 2. Measuring regularity

In this section, we first introduce four measures that can be used to quantify the degree of regularity of polysemy patterns and then use semantic information extracted from the WordNet corpus to compute these four measures for 15 polysemy patterns in English. These patterns were selected based on the previous research on polysemy (Apresjan, 1974; Brocher et al. 2016, 2018; Carston & Wearing, 2011; Klepousniotou, 2002; Klepousniotou et al., 2012; Lakoff & Johnson, 1980; Srinivasan & Rabagliati, 2015) as well as our professional knowledge.

### 2.1 Measures of regularity

Our work builds on a previous verbal description of how polysemy regularity might be derived from corpora (Lombard et al., 2023). Lombard et al. (2023) proposed that a given pattern’s regularity could be quantified as a ratio of two quantities. The numerator represents the number of words that instantiate a given pattern: that is, the number of words that, in addition to their “base” sense (SENSE 1), also have a “target” sense created through the process of sense extension

(SENSE 2). In turn, the denominator is the number of words that have SENSE 1 (irrespective of whether they instantiate the pattern). This ratio can be calculated either in a type-based or in a token-based manner. However, because no lexical-semantic database such as WordNet is currently available for French, Lombard et al. (2023) were unable to extract the information necessary to actually compute the proposed measures. The study reported here extends Lombard et al.’s proposal to quantify polysemy patterns in English.

Four measures were developed to capture to what extent words that have one sense also have another sense derived through a polysemic extension (e.g., SENSE 1 of *pig* is ‘animal’ and SENSE 2 is ‘person’). Measure (1),  $R_1$ , is a count of the number of words that have both SENSE 1 and SENSE 2 in a given pattern. Measure (2),  $R_2$ , is a ratio of  $R_1$  and the number of words with SENSE 1, regardless of whether they also have SENSE 2. Thus, while  $R_1$  represents a raw count of the number of words that instantiate a pattern,  $R_2$  reflects the instantiation rate of the pattern amongst the words on which this pattern could in theory operate. In this respect,  $R_2$  is comparable to the measure of morphological productivity proposed by Aronoff (1976) (the ratio of all attested words to all possible words for a given morphological pattern). In (1) and (2) below,  $N_{s_2}$  denotes the number of words that have both senses, and in (2),  $N_{s_1}$  represents the number of words with at least SENSE 1.

$$(1) \quad R_1 = N_{s_2}$$

$$(2) \quad R_2 = \frac{N_{s_2}}{N_{s_1}}$$

To account for the fact that regularity is likely to play a more important role when a pattern is instantiated by a very frequent as opposed to an infrequent word, regularity measures  $R_3$  and  $R_4$  (shown below in (3) and (4), respectively) weight the first two regularity measures,  $R_1$  and  $R_2$ , by the log-frequency of occurrence of the word. In formulae given in (3) and (4) below,  $f_w$  represents the form frequency of a given word ( $w$ ).

$$(3) \quad R_3 = \sum_{w=1}^{N_{s_2}} \log(f_w)$$

$$(4) \quad R_4 = \frac{\sum_{w=1}^{N_{s_2}} \log(f_w)}{\sum_{w=1}^{N_{s_1}} \log(f_w)}$$

## 2.2 Pattern extraction from corpora

We used WordNet (Fellbaum, 1998; Miller, 2005) to identify words with one or two senses for each of the polysemy patterns under investigation (see **Table 1**). WordNet is an English lexical database that contains semantic structures for 117,000 *synsets*, where a synset is defined as

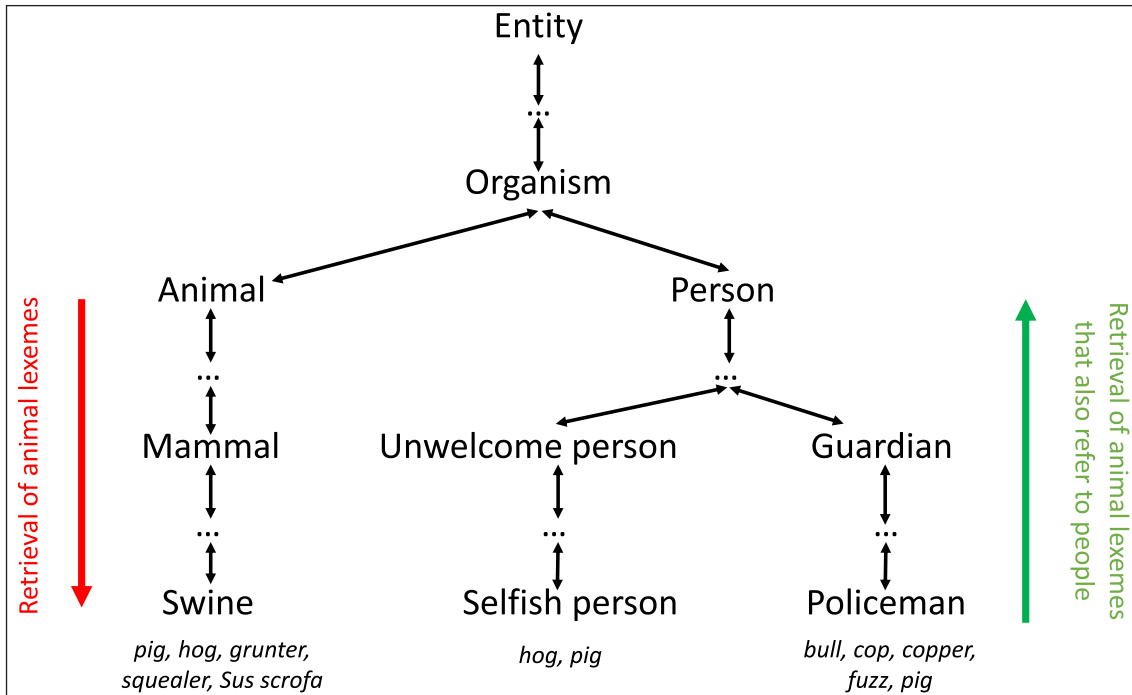
an unordered set of synonyms organised around one sense (i.e., words that denote the same concept and are interchangeable in many contexts). For example, *pig* has 6 senses in WordNet and thus appears in 6 different groups of synonyms (see example (5) below), where each group is linked to others through some semantic relationship. One possible type of semantic relationship is hyponymy (also called *hypernymy* or *super-subordinate relation*): the association between a specific term, a *hyponym*, and a generic term, a *hypernym*, that includes all the semantic features of the hyponym (e.g., *pig* and *animal*). Words that belong to the same level in the hierarchy are called co-hyponyms (e.g., *pig*, *dog*, *mouse*, *cat*, etc.). In WordNet, each word (e.g., *pig*) appears in a hierarchical tree that links it to its more specific subtype sense (e.g., *porker*) as well as to its broader senses up to a root node (e.g., *swine*, *even-toed ungulate*, *placental mammal*, *mammal*, *vertebrate*, *chordate*, *animal*, *organism*, *living thing*, *whole*, *physical object*, *physical entity*, *entity*).

- (5)
- a. *hog, pig, grunter, squealer, Sus scrofa* — ‘domestic swine’
  - b. *slob, sloven, pig, slovenly person* — ‘a coarse obnoxious person’
  - c. *hog, pig* — ‘a person regarded as greedy and pig-like’
  - d. *bull, cop, copper, fuzz, pig* — ‘uncomplimentary terms for a policeman’
  - e. *pig bed, pig* — ‘mold consisting of a bend of sand in which pig iron is cast’
  - f. *pig* — ‘a crude block of metal (lead or iron) poured from a smelting furnace’

This hierarchical organisation of the different groups of synonyms in WordNet makes it possible to determine the degree of regularity for each polysemy pattern. More specifically, for each polysemy pattern SENSE 1 → SENSE 2, one can retrieve all synonyms of all words that have SENSE 1 as their hypernym (e.g., ‘animal’). Then, for each of these synonyms, it can be verified whether it also has a sense for which SENSE 2 is a hypernym (e.g., ‘person’). **Figure 1** illustrates with the example of *pig* how the hierarchical tree structure of WordNet can be used to determine how many words with SENSE 1 (‘animal’) also have SENSE 2 (‘person’).

We performed the steps described above separately for 15 polysemy patterns using the NLTK library (Bird et al., 2009) in Python ([www.python.org](http://www.python.org)). By selecting these specific patterns, we attempted to sample as many unambiguous metaphor patterns as possible, within the limits of our linguistic knowledge. For each of these patterns, the procedure discussed above resulted in a list of all words with SENSE 1 and an indication of whether each of these words also had SENSE 2. For example, for the ANIMAL → PERSON pattern, the words and expressions in (6a) have both senses, while those in (6b) do not.

- (6)
- a. *hog, pig, dog, bear, snake, black sheep, guinea pig, killer bee, water rat, ...*
  - b. *grunter, Sus scrofa, fish, allosaurus, angler fish, angoumois moth, ...*



**Figure 1:** Hypernym structure of  $pig_a$ ,  $pig_c$ , and  $pig_d$  from (5).

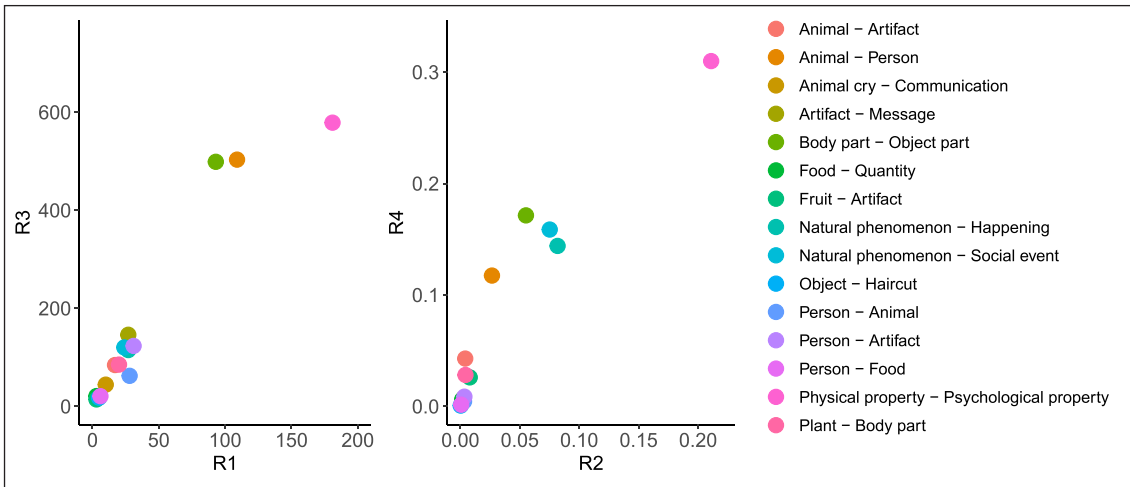
Next, for every word contained in the lists, we retrieved the frequency of occurrence as reported in Subtlex-UK (Van Heuven et al., 2005). Because Subtlex-UK contains only unigrams, we had to exclude all multiword expressions (e.g., *Sus scrofa*, *black sheep*, *guinea pig*, *killer bee*, *water rat*) and sequences that were not part of Subtlex-UK (e.g., *geometry teacher*, *ungulated animal*, or *bad person*) from our lists. Furthermore, we also excluded all words that were not listed in the Oxford English Dictionary (OED Online, 2022), which resulted in removal of 28.220% of our original sample of 103,219 words. We then manually verified that each remaining word had both SENSE 1 and SENSE 2 and further removed the polysemous words whose semantic extension was not based on metaphor. By the end of pre-processing, our sample consisted of 74,093 unique lexemes, and we then computed the four regularity measures for each of the 15 patterns (**Table 1**). Pearson correlation tests (see **Table 2**) showed that there are strong positive correlations between the corresponding type-based and frequency-weighted measures for both the count ( $R_1$  and  $R_3$ ) and the ratio ( $R_2$  and  $R_4$ ) measures. **Figure 2** depicts the relationship between the four regularity measures for each of the 15 patterns. The technical information about the extraction and pre-processing of polysemy patterns are available on the project's repository on OSF: <https://osf.io/uhy75/>.



**Table 1:** Regularity measures. For each pattern, the table reports the number of words with both SENSE 1 and 2 (S1&S2) and the number of words with SENSE 1 (S1) only. Patterns chosen for the experimental part of our study are indicated with an asterisk (\*).

Pattern (SENSE 2 → SENSE 1)	S1&S2	S1	R1	R2	R3	R4	Examples
*ANIMAL → ARTIFACT	17	4101	17	0.004	84.298	0.043	<i>mouse, bug, drone, ...</i>
*ANIMAL → PERSON	109	3989	109	0.027	503.309	0.117	<i>pig, shark, fox, ...</i>
ANIMAL CRY → COMMUNICATION	10	27	10	0.270	44.283	0.442	<i>bark, cackle, howl, ...</i>
*ARTIFACT → MESSAGE	27	11888	27	0.002	146.063	0.007	<i>garbage, rubbish, ...</i>
*BODY PART → OBJECT PART	93	1591	93	0.055	498.860	0.171	<i>heart, head, leg, ...</i>
FOOD → QUANTITY	3	2010	3	0.002	21.009	0.006	<i>mess, grain, cud</i>
FRUIT → ARTIFACT	3	379	3	0.008	14.142	0.026	<i>grape, pod, nut</i>
*NATURAL EVENT → HAPPENING	27	303	27	0.082	114.825	0.144	<i>flood, rain, wave, ...</i>
NATURAL EVENT → SOCIAL EVENT	24	295	24	0.075	120.208	0.159	<i>wind, earthquake, ...</i>
OBJECT → HAIRCUT	5	16507	5	0.0003	17.720	0.001	<i>beehive, thatch, ...</i>
*PERSON → ANIMAL	28	8967	28	0.003	62.302	0.004	<i>queen, emperor, ...</i>
PERSON → ARTIFACT	31	8961	31	0.003	123.299	0.009	<i>secretary, host, ...</i>
PERSON → FOOD	6	8986	6	0.001	20.766	0.002	<i>bomber, marquise, ...</i>
*PHYSICAL PROPERTY → PSYCHOLOGICAL PROPERTY	181	677	181	0.211	578.707	0.310	<i>depth, acidity, ...</i>
PLANT → BODY PART	20	4620	20	0.004	85.220	0.028	<i>bulb, bush, iris, ...</i>





**Figure 2:** Relationship between the four regularity measures for each pattern:  $R_1$  and  $R_3$  (count-based measures) on the left and  $R_2$  and  $R_4$  (ratio-based measures) on the right.

**Table 2:** Data of pairwise Pearson correlation test between the metrics.

	R1	R2	R3	R4
R1	–	–	–	–
R2	$r = .416, p = .123$	–	–	–
R3	$r = .958, p < .001$	$r = .327, p = .234$	–	–
R4	$r = .454, p = .089$	$r = .975, p < .001$	$r = .421, p = .119$	–

### 3. Experimental study

We conducted an online experiment to assess (a) whether adults have sufficient knowledge of the polysemy patterns to be able to generalise from them; and (b) whether this knowledge is graded according to the different metrics of regularity described above. Our experiment was based on the notion that adults' knowledge of a polysemy pattern should modulate the extent to which they accept new sense extensions from that pattern. When patterns are highly regular, adults should be more likely to accept new sense extensions than for less regular patterns. To test this hypothesis, we developed a semantic judgement task with semantic neologisms as experimental stimuli. These neologisms were existing words that were used in a new context in our experiment, thus in a different sense, such as *the moustache of the broom* for *moustache* (Bastuji, 1974; Renouf, 2013; Smyk-Bhattacharjee, 2009). We then measured the extent to which

adults found the context sentences plausible. In the remainder of this section, we describe our stimuli, participants, and procedure. Materials, data, and analysis code for the experimental work are available at <https://osf.io/uhy75/>.

## 3.1 Methods

### 3.1.1 Materials

We selected seven patterns from the fifteen listed in **Table 1** (indicated with an asterisk) to ensure a wide range of regularity. In selecting patterns for the experimental study, we kept only those patterns for which we were certain that new senses could be created. For example, the pattern ANIMAL CRY → COMMUNICATION (e.g., *roar*, *bark*, *cackle*) was not included because it is instantiated by so many words that it is difficult to create neologisms.

For each of the seven patterns, we created five semantic neologisms that followed the semantic extension characterising the pattern. In (7a) and (8a) below, we provide examples of words that we used to create the semantic neologisms. These words have SENSE 1 but not SENSE 2; however, there is no reason why they could not develop SENSE 2. Therefore, we created semantic neologisms by inserting these words into sentence frames intended to realise SENSE 2 (see **Table 2** for examples). Hereafter, these semantic neologisms are referred to as ‘new’ items; there were 35 ‘new’ items in total.

The acceptability of these ‘new’ items in sentence contexts was compared to the acceptability of semantic neologisms that were not part of the pattern, in that they did not take SENSE 1. These ‘illegal’ items (in (7b) and (8b) in examples below) were matched groupwise to the target ‘new’ items (in (7a) and (8a) below) for frequency of occurrence and number of letters. ‘New’ items and ‘illegal’ items were all nouns. In the experiment, like the ‘new’ items, the ‘illegal’ items were presented in sentence contexts intended to realise SENSE 2 (see **Table 2** for examples).

- (7) ANIMAL → OBJECT
- a. New: *chicken*, *crab*, *porcupine*
  - b. Illegal: *grandpa*, *tutor*, *cousin*
- (8) BODY PART → OBJECT PART
- a. New: *bone*, *antler*, *knee*
  - b. Illegal: *salad*, *curry*, *milk*

Among our patterns, only PHYSICAL PROPERTY → PSYCHOLOGICAL PROPERTY had an abstract SENSE 1. Because concreteness<sup>1</sup> may have an effect on participants’ acceptability judgements,

---

<sup>1</sup> There are many possible definitions of concreteness (see, e.g., Barsalou, 2003; Huyghe, 2015; Kleiber & Vuillaume, 2011; Van de Velde, 1995). Here we distinguish concrete and abstract meanings based on whether their referent can be considered a physical object.

we decided to pair the ‘new’ items for this pattern with two types of ‘illegal’ items, one with an abstract SENSE 1 ( $N = 5$ ) and one with a concrete SENSE 1 ( $N = 5$ ). Thus, there were 10 ‘illegal’ items for this pattern, but 5 ‘illegal’ items for all other patterns, such that the total number of ‘illegal’ items was 40.

Sentence frames used for the ‘new’ and ‘illegal’ items were closely matched on their syntactic structure (e.g., *I think that the [Target] needs more [Noun] to [Verb].*). To ensure that the effect of interest was not contaminated by effects arising due to particular combinations of new senses and sentence frames, a second version of the experimental stimuli was created. In this version, the targets were rotated into different sentence frames chosen from the same set as in the first version. Stimuli version was counterbalanced across participants, with an approximately equal number of participants per version (see 3.1.2).

Finally, to avoid a situation in which all sentences in the experiment contained a semantic neologism, we added 40 existing metaphors in appropriate sentence contexts as fillers (note that the filler items were identical across the two versions of the stimuli). These filler items were not matched to ‘new’ and ‘illegal’ items for polysemy pattern, word length, or word frequency, and hence were not included in the statistical analyses. Examples of filler items are provided in **Table 3**. Overall, each participant was exposed to a total of 115 sentence contexts (35 ‘new’, 40 ‘illegal’, and 40 fillers).

**Table 3:** Examples of the experimental and filler items.

Pattern	Condition	Word	Sentence
ANIMAL → ARTIFACT	New	<i>chicken</i>	I think that the <b>chicken</b> needs more oil to function.
	Illegal	<i>grandpa</i>	I think that the <b>grandpa</b> needs more fuel to start.
BODY PART → OBJECT PART	New	<i>knee</i>	I always slow down on the <b>knee</b> of the mountain road.
	Illegal	<i>milk</i>	I always meditate on the <b>milk</b> of the highest hill.
ANIMAL → ARTIFACT	Filler	<i>mouse</i>	I think that the <b>mouse</b> needs more charge to work
ARTIFACT → ARTIFACT	Filler	<i>hammer</i>	I’ve never pressed on the <b>hammer</b> of my shotgun.

### 3.1.2 Participants

Thirty-two students aged between 18 and 30 (Mean = 24.906, SD = 6.326) took part in the experiment. All participants reported English as their first language. The participants were recruited through the Prolific crowd-sourcing platform and reimbursed with £3 for their time (i.e., £12/hour). Participants were randomly assigned to one of the two versions of the experimental materials (17 and 15 participants, respectively). The number of participants reflected the resources available. There was insufficient prior work to specify parameters for an *a priori* power analysis.

### 3.1.3 Procedure

The experiment took the form of a 15-minute online survey. Each item was presented in two steps: first, with a hyphen in place of the target word (9a) and then with the target word (in bold) filling that gap (9b). The sentences were presented using the default parameters (Open Sans, 22pt) of the survey platform that we used (Qualtrics). To make it easier for the participants to focus on the target word, each sentence was split across three lines, with the gap/target word presented on the second line and the text preceding and following the gap/target word on the first and third lines, respectively. For each step, the participants could take as much time as they needed to read the sentence, and each sentence remained on the screen until the participants pressed a key indicating that they were ready to continue.

- (9) a. She said that the – of his smile really charmed her.  
 g. She said that the **warmth** of his smile really charmed her.

The participants were asked to judge the plausibility of the sentences presented. The instructions were as follows: “You will see a sentence with a gap and then the same sentence with a word filling that gap. Please, read the sentence carefully and then decide how plausible it seems to you. Indicate this with the help of the cursor; left extreme – ‘no sense at all’, right extreme – ‘completely acceptable.’”

Participants judged the plausibility of the sentences using a scale ranging from 0 to 100, where 0 stood for ‘no sense at all’ and 100 stood for ‘completely acceptable’. The scale was represented by a horizontal bar at the bottom of the screen, and the participants were told to indicate their response by sliding the cursor on that bar. The instructions did not specify whether the participants were to judge the semantic properties of the target words, but, as stated above, by presenting each sentence across three lines, we attempted to draw the participants’ attention to the target words.

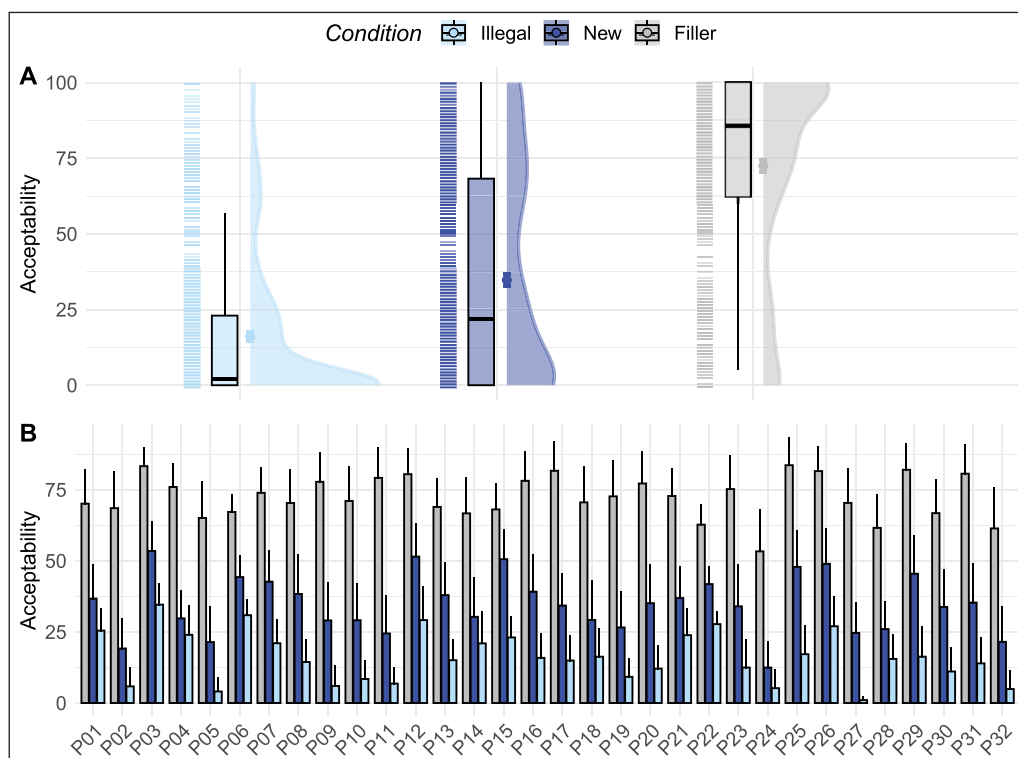
## 3.2 Results

In line with the two research aims stated previously, two analyses were conducted. First, we analysed the effect of pattern knowledge: the independent variable was condition, i.e., whether

the semantic neologism was part of a pattern (i.e., ‘new’ items) or not (i.e., ‘illegal’ items). Second, we analysed the effect of pattern regularity; in this analysis, the independent variable was the regularity of the pattern being instantiated. This variable was operationalised using the four regularity metrics ( $R_1$ – $R_4$ ) described in Section 2, and we also examined which regularity measure could best account for the data. Our hypotheses were (a) that participants would rate ‘new’ senses as more acceptable than ‘illegal’ senses; and (b) that the acceptability of ‘new’ senses would increase as a function of pattern regularity.

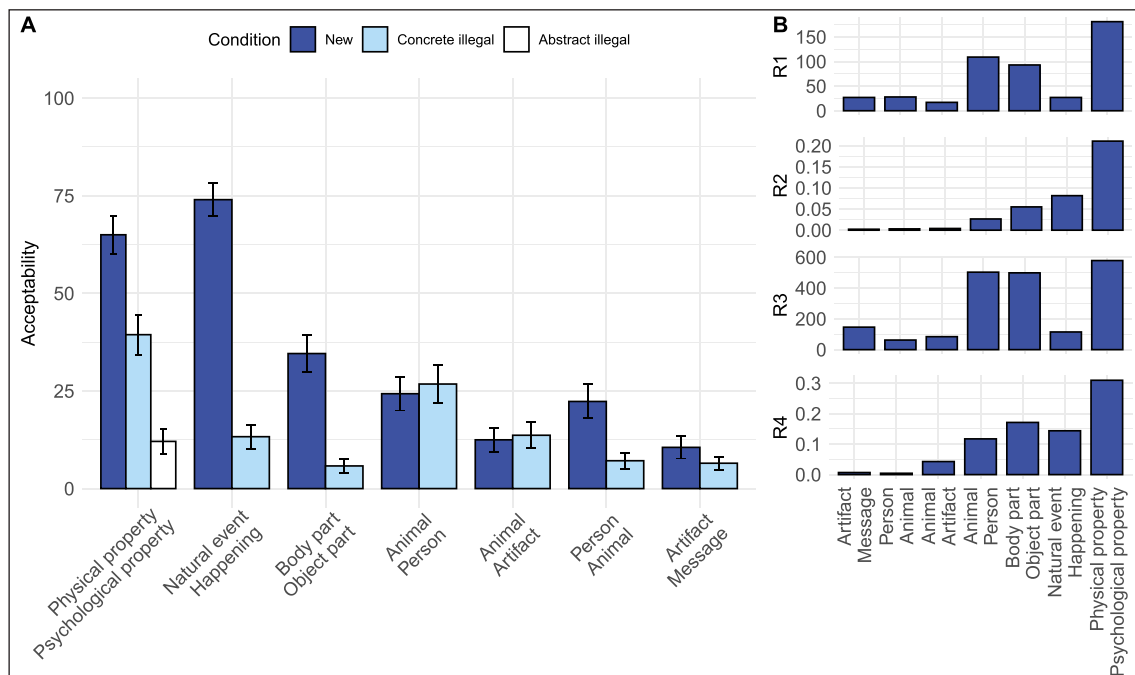
### 3.2.1 Analysis 1: Effect of pattern knowledge on plausibility ratings

**Figure 3a** illustrates the distribution of the acceptability ratings for ‘new’, ‘illegal’, and filler items. **Figure 3b** shows that, although the participants varied in their acceptability ratings, for each condition, there was a high degree of consistency in their ratings, with high ratings for existing filler items (Mean = 72.468, SD = 33.784), medium-high ratings for ‘new’ items (Mean = 34.750, SD = 35.352), and low ratings for ‘illegal’ items (Mean = 16.080, SD = 24.566).



**Figure 3:** Acceptability ratings across the three conditions. Panel A shows the overall distribution of ratings (“rain clouds”; with means and standard errors for each condition) and the interquartile range of the data (“box-and-whiskers”; medians are represented by black horizontal lines). The barcode represents individual data points (with darker colours showing where more data points cluster) for each condition. Panel B shows the mean ratings per participant and condition. Error bars correspond to 2 standard errors from the mean.

The average acceptability ratings for each pattern for the ‘new’ and ‘illegal’ conditions are shown in **Figure 4**, and a figure depicting the acceptability of each item (grouped by patterns) is available on the OSF site for this project. Descriptively, the most regular patterns (based on  $R_2$ , on the left) showed a large difference between ‘new’ and ‘illegal’ senses, whereas the less regular patterns (on the right) showed a smaller difference. The large numerical difference between the two ‘illegal’ conditions (abstract and concrete) linked to the pattern PHYSICAL PROPERTY → PSYCHOLOGICAL PROPERTY could indicate that concreteness of new senses may impact the extent to which speakers consider them acceptable, with the new abstract senses being more likely to be judged acceptable than the new concrete senses. However, this observation will need to be examined empirically in future studies. In the analyses reported below, only the data from the concrete ‘illegal’ condition for this pattern were considered, as this provided the most conservative test of the difference between ‘new’ and ‘illegal’ items.



**Figure 4:** Panel A shows the average acceptability ratings per polysemy pattern and experimental condition. Error bars correspond to 2 standard errors from the mean. Patterns are ordered by regularity according to  $R_2$ . Panel B shows the regularity value per polysemy pattern for each regularity metric ( $R_1$ – $R_4$ ).

In Analysis 1, the data were analysed using a linear mixed effects model with *condition* as a fixed effect, which was sum-coded, and random intercepts for participants and items as random effects. This model was obtained following the approach proposed in Bates et al.

(2018), which uses the principal component analysis to decide on the most parsimonious model (i.e., a model with the maximal – in the frequentist framework – random-effects structure that is supported by both the design and the data). In brief, we started with a model that also included the by-subject adjustments to the random slopes for condition (as this was the maximal random-effects structure that was sensible given the design of our study); however, because this component only accounted for 0.3% of explained variance, it was removed. The resulting model showed that ‘new’ items were judged as more acceptable than the ‘illegal’ items ( $\beta = 18.670$ ,  $CI = [7.478, 23.349]$ ,  $p < .001$ ,  $AIC: 20,364.35$ , 2,240 observations).

### 3.2.2 Analysis 2: Effect of pattern regularity on plausibility ratings

The effect of regularity was assessed through linear mixed effects models, where only the data for the ‘new’ items ( $N = 35$ ; 1,120 observations in total) were included. We excluded ‘illegal’ items because they contained neologisms that did not follow any pattern and thus could not be characterised in terms of regularity.

We ran four linear mixed effects models, with one model per regularity measure (as a continuous independent variable). As in Analysis 1, we followed the procedure described in Bates et al. (2018) to determine the most parsimonious model given the the design and the data, with the final model including random intercepts for participants and items. As in Analysis 1, we always started with models that included by-subject adjustments to the random slope for regularity; however, only in models for  $R_2$  and  $R_4$  did this component account for some of the explained variance (5% and 12%, respectively). In models for  $R_1$  and  $R_3$ , the random slope adjustment did not account for any variance over and above that captured by the by-subject adjustment to the intercept and was therefore removed. As in Analysis 1, we used log-likelihood tests to determine whether the null hypothesis of no effect could be rejected. The predicted slope for each model is shown in **Figure 5**. The results (see **Table 4**) indicated that all measures of regularity except for  $R_3$  were significant predictors of the acceptability judgements for the ‘new’ senses. A comparison of the Akaike Information Criterion (AIC; Akaike, 1973) and the Bayesian Information Criterion (BIC; Schwarz, 1978) suggested that  $R_2^2$  and  $R_4$  were better predictors of acceptability judgements than  $R_1$  and  $R_3$ .

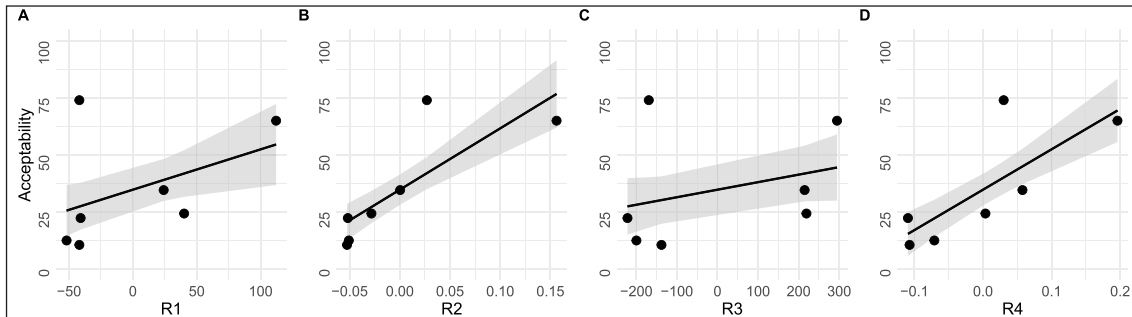
---

<sup>2</sup> One reviewer pointed out that  $\log(R_2)$  replicates all results but with a higher AIC than that for the  $R_2$  model. This observation suggests that  $\log(R_2)$  may be a measure of regularity superior to  $R_2$ . However, we have retained  $R_2$ , because it was the measure that we set out to test (based on earlier work, see Lombard et al., 2023), and because it links to Aronoff’s (1976) theory of productivity in morphology. To facilitate comparison of these two measures, we have included analysis code for both  $R_2$  and  $\log(R_2)$  in our OSF repository.



**Table 4:** Results of the regularity analysis.

	<i>AIC</i>	<i>BIC</i>	$\beta$	SE	t-value	<i>p</i>
R <sub>1</sub>	10419.270	10444.380	0.177	0.071	2.494	.014
R <sub>2</sub>	10383.500	10413.630	268.916	43.050	6.247	<.001
R <sub>3</sub>	10424.950	10450.050	0.033	0.020	1.666	.090
R <sub>4</sub>	10389.280	10419.410	177.102	31.157	5.684	<.001

**Figure 5:** Predicted acceptability ratings extracted from the linear mixed effects models for each of the four regularity metrics (Panel A: R<sub>1</sub>, Panel B: R<sub>2</sub>, Panel C: R<sub>3</sub>, Panel D: R<sub>4</sub>). The black dots represent the observed mean acceptability ratings for each of the seven patterns.

## 4. General discussion

Although polysemy is a fundamental characteristic of language, it remains poorly understood. Linguists have often observed that sense extensions in polysemy follow patterns (Apresjan, 1974; Barque et al., 2018; Copestake & Briscoe, 1995; Dölling, 2020; Nunberg, 1995; Nunberg & Zaenen, 1992; Pustejovsky, 1995); however, to date, the regularity of these patterns has not been well characterised. In psycholinguistics, polysemy is often described in a categorical manner; for example, a word is described as ambiguous or unambiguous (e.g., Rodd et al., 2002) or as regular or irregular (e.g., Rabagliati & Snedeker, 2013). Likewise, because metonymy is often assumed to be more regular than metaphor (Apresjan, 1974), some researchers have associated regular polysemy with metonymy and irregular polysemy with metaphor (Brocher et al., 2018; Klepousniotou et al., 2012).

The results of our regularity analysis demonstrate that it is possible to quantify the regularity of English polysemy patterns in a graded manner. The wide variation in regularity highlighted in our metrics is inconsistent with a characterisation of metaphor as ‘irregular’ (cf. Brocher

et al., 2018; Klepousniotou et al., 2012). Although some of our metaphor patterns had very low regularity, these were at one end of a continuum, with other patterns being of such high regularity that it was difficult to find non-existing sense extensions. This initial investigation does not allow us to comment on how metaphor and metonymy compare in terms of the regularity of the patterns that they instantiate, but it would be possible to conduct this investigation using the methods and metrics that we have developed. Likewise, while our study was limited to English, it should be possible to apply our methods to other languages to determine the extent to which particular patterns are more or less regular across languages. This work would enhance the robustness of Srinivasan and Rabagliati (2015)'s initial attempts to measure cross-linguistic regularity.

Our approach to quantifying polysemy regularity is similar to attempts to quantifying productivity in morphology. The concept of productivity is usually decomposed into availability (categorical) and profitability (numeric) (Bauer, 2001). A morphological process is available if it can be used to coin new words in a certain situation. For instance, *-ness* can form novel deadjectival nouns in contemporary English and is thus available, in contrast to *-th*. Profitability refers to how frequently an available process is used to coin new derivatives, either in the past or in the present production. Profitability can be based either on the number of existing words that are formed through a morphological process, or on the propensity of a morphological process to create neologisms synchronically. Thus, measures of productivity based on profitability have some similarities to our metrics of polysemy regularity. For example, the size of the morphological series is close to our  $R_1$  measure, and Aronoff (1976)'s definition of productivity as a ratio between attested and theoretically possible words is close to our  $R_2$  measure. Other conceptions of productivity are further apart, since they aim to measure the propensity of a morphological process to create new words rather than its rate of instantiation in a language (e.g., Baayen, 1992, 1994, 2009; Lindsay & Aronoff, 2013; Plag, 1999; Spencer, 2019). Equivalent measures of polysemy productivity (i.e., the extent to which each polysemy pattern is important in the renewal of the lexicon) could be determined in a future study.

Importantly, our research findings go beyond a new linguistic description. Our data show that our characterisation of regularity can predict human acceptability judgements of novel sense extensions. In our experiment, novel sense extensions based on more regular patterns were judged as more acceptable than those based on less regular patterns. Our interpretation of this finding is that language users' experience with these patterns shapes their intuition regarding the acceptability of novel senses. The fact that this result holds particularly for metrics  $R_2$  and  $R_4$  suggests that it is not the number of times that a pattern is instantiated that is important, but rather the consistency with which words that have SENSE 1 also have SENSE 2. Our results do not allow us to differentiate between these two metrics ( $R_2$  and  $R_4$ ), because they are highly correlated for

the patterns studied in our analyses (**Figure 2**); however, distinguishing whether regularity is best measured using a type-based or a frequency-weighted ratio could be an important avenue for future research.

Our data further show that speakers tend to link two senses with each other more easily when these senses have a high probability of association (i.e., when both  $R_2$  and  $R_4$  are high). One possible reason for this finding could be that polysemy patterns, like individual polysemes, are represented in the mental lexicon, and that the strength of their representations is modulated by pattern consistency. At the level of individual words, if many words exhibit a particular semantic relationship, words that share the same semantic relationship between their two senses could have a type of parallel organisation in the mental lexicon that favours rule-based semantic analysis. The presence of such representations could explain why speakers generalise some polysemy patterns more readily than others, and it would be worthwhile to test this hypothesis in the future. One might also draw a parallel with the dual-route models in morphology (Bertram et al., 2000; Pinker, 1999; Pinker & Prince, 1988), and consider whether the different polysemous meanings are learnt and stored, or whether they are inferred from the patterns of polysemy, or both.

It is important to recognise the limitations of this research. Our study was intended to provide initial evidence regarding the objective measurement of regularity of English polysemy patterns. However, our conclusions are necessarily limited by the small number of polysemy patterns included in the acceptability study. This limitation means that we cannot definitively determine whether the effect of regularity on acceptability judgements is scalar or categorical. Likewise, the strong positive relationships between type-based and frequency-weighted regularity metrics in this limited sample of polysemy patterns means that we cannot determine which of these metrics provides a superior description of regularity. It will be important for future studies to replicate our findings with more polysemy patterns that vary widely in terms of regularity. Similarly, while this study provides initial evidence for the psychological reality of polysemy regularity, pre-registered confirmatory studies are needed to establish this phenomenon more robustly.

Future research should also investigate the potential effect of concreteness on acceptability judgements. Because one of our patterns (PHYSICAL PROPERTY → PSYCHOLOGICAL PROPERTY) was characterised by an abstract SENSE 2, we created two types of ‘illegal’ sense extensions, one that was concrete (to match the other ‘illegal’ sense extensions) and one that was abstract (to match the SENSE 2 extensions for this pattern). Since we did not have a prediction about how concreteness might influence acceptability judgements, the observed results need to be treated with caution. However, there is an indication that abstract sense extensions may be judged as more acceptable than concrete sense extensions. If confirmed empirically, this observation may have important methodological implications in terms of the need for investigators to control

the concreteness of foils in experiments of this nature. Furthermore, this issue raises interesting theoretical questions about why abstract words may be better able to take on new meanings than concrete words.

In summary, the research presented here has pioneered an approach for quantifying the regularity of polysemy patterns in a graded manner. We have demonstrated that this approach has psychological validity, and we hope that this work inspires new work investigating the influence of regularity on the learning, processing, and storage of polysemes in a more precise and nuanced way than has previously been possible under categorical conceptualisations. The approach documented here could be used to study how the consistency of polysemy patterns changes through the course of language development, and also how polysemy patterns vary across the world's languages.

---

## Data accessibility statement

Material, data, pattern extraction code, and analysis code are available on OSF: <https://osf.io/uhy75/>.

## Acknowledgements

We are grateful to the two anonymous reviewers and Dr. João Veríssimo for their valuable comments and suggestions on an earlier version of this paper.

## Funding information

This work was conducted while the first author was a visiting PhD student at Royal Holloway, University of London, supported by a Doc.Mobility grant (DM-21-03) from the Research Promotion Committee of the University of Fribourg. This research was further supported by research grants to KR from the Economic and Social Research Council (ES/W002310/1) and the Leverhulme Trust (RPG-2020-034).

## Competing interests

The authors have no competing interests to declare.

## Authors' contributions

AL, AU, and KR conceived the project and designed the experiment; AL and AU extracted polysemy pattern information and defined the regularity measures; AL collected the experimental data; AL and MK analysed the experimental data; AL, AU, MK and KR wrote and edited the manuscript; KR supervised the project.

---

## References

- (2022). *OED Online*. Oxford University Press. <https://www.oed.com>. Accessed in March 2022.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory*, 267–281. DOI: [https://doi.org/10.1007/978-1-4612-1694-0\\_15](https://doi.org/10.1007/978-1-4612-1694-0_15)
- Apresjan, J. (1974). Regular polysemy. *Linguistics*, 42, 5–32. DOI: <https://doi.org/10.1515/ling.1974.12.142.5>
- Aronoff, M. (1976). *Word formation in generative grammar*. MIT Press.
- Baayen, R. H. (1992). Statistical models for word frequency distribution: A linguistic evaluation. *Computers and Humanities*, 26(5/6), 347–363. DOI: <https://doi.org/10.1007/BF00136980>

- Baayen, R. H. (1994). Derivational productivity and text typology. *Journal of Quantitative Linguistics*, 1(1), 16–34. DOI: <https://doi.org/10.1080/09296179408589996>
- Baayen, R. H. (2009). Corpus linguistics in morphology: Morphological productivity. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international Handbook*, 899–919. De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110213881.2.899>
- Barque, L., Haas, P., & Huyghe, R. (2018). Polysémie régulière et néologie sémantique. Constitution d'une ressource pour l'étude des sens nouveaux. *Langages*, 12, 91–108. DOI: <https://doi.org/10.15122/isbn.978-2-406-08196-8.p.0091>
- Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society London*, 358, 1177–1187. DOI: <https://doi.org/10.1098/rstb>
- Bastuji, J. (1974). Aspects de la néologie sémantique. *Langages*, 36, 6–19. DOI: <https://doi.org/10.3406/lgge.1974.2270>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2018). Parsimonious mixed models. *arXiv*, 1506. DOI: <https://doi.org/10.48550/arXiv.1506.04967>
- Bauer, L. (2001). *Morphological productivity*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511486210>
- Bertram, R., Schreuder, R., & Baayen, R. H. (2000). The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and productivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(2), 489–511. DOI: <https://doi.org/10.1037/0278-7393.26.2.489>
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python*. O'Reilly Media Inc. <https://www.nltk.org/book/>.
- Brocher, A., Koenig, J.-P., & Foraker, S. (2016). Processing of irregular polysemes in sentence reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 42(11), 1798–1813. DOI: <https://doi.org/10.1037/xlm0000271>
- Brocher, A., Koenig, J.-P., Mauner, G., & Foraker, S. (2018). About sharing and commitment: The retrieval of biased and balanced irregular polysemes. *Language, Cognition and Neuroscience*, 33(4), 443–466. DOI: <https://doi.org/10.1080/23273798.2017.1381748>
- Carston, R., & Wearing, C. (2011). Metaphor, hyperbole and simile: A pragmatic approach. *Language and Cognition*, 3, 283–312. DOI: <https://doi.org/10.1515/langcog.2011.010>
- Copetake, A., & Briscoe, T. (1995). Semi-productive polysemy and sense extension. *Journal of Semantics*, 12(1), 15–67. DOI: <https://doi.org/10.1093/jos/12.1.15>
- Dölling, J. (2020). Systematic Polysemy. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann & T. E. Zimmermann (Eds.), *The Wiley Blackwell companion to semantics*. DOI: <https://doi.org/10.1002/9781118788516.sem099>
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27, 429–446. DOI: [https://doi.org/10.1016/0749-596X\(88\)90066-6](https://doi.org/10.1016/0749-596X(88)90066-6)

- Eddington, C. M., & Tokowicz, N. (2015). How meaning similarity influences ambiguous word processing: The current state of the literature. *Psychonomic Bulletin and Review*, 22(1), 13–37. DOI: <https://doi.org/10.3758/s13423-014-0665-7>
- Falkum, I. L., & Vicente, A. (2015). Polysemy: Current perspectives and approaches. *Lingua* 157, 1–16. DOI: <https://doi.org/10.1016/j.lingua.2015.02.002>
- Fang, X., Perfetti, C., & Stafura, J. (2017). Learning new meanings for known words: Biphasic effects of prior knowledge. *Language, Cognition and Neuroscience* 32(5), 637–649. DOI: <https://doi.org/10.1080/23273798.2016.1252050>
- Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. MIT Press. DOI: <https://doi.org/10.7551/mitpress/7287.001.0001>
- Filipović Đurđević, D., & Kostić, A. (2023). We probably sense sense probabilities. *Language, Cognition and Neuroscience*, 38(4), 471–498. DOI: <https://doi.org/10.1080/23273798.2021.1909083>
- Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, 29, 181–200. DOI: [https://doi.org/10.1016/0749-596X\(90\)90071-7](https://doi.org/10.1016/0749-596X(90)90071-7)
- Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(6), 1366–1383. DOI: <https://doi.org/10.1037//0278-7393.25.6.1366>
- Frisson, S. (2009). Semantic underspecification in language processing. *Language and Linguistics Compass*, 3(1), 111–127. DOI: <https://doi.org/10.1111/j.1749-818X.2008.00104.x>
- Gibson, E., Futrell, R., Piantadosi, S.P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, 23(5), 389–407. DOI: <https://doi.org/10.1016/j.tics.2019.02.003>
- Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, 12, 395–427. DOI: <https://doi.org/10.3758/BF03201693>
- Huyghe, R. (2015). Les typologies nominales: présentation. *Langue Française*, 185, 5–27. DOI: <https://doi.org/10.3917/lf.185.0005>
- Johnson-Laird, P. N. (1987). The mental representation of the meaning of words. *Cognition*, 25(1–2), 189–211. DOI: [https://doi.org/10.1016/0010-0277\(87\)90009-6](https://doi.org/10.1016/0010-0277(87)90009-6)
- Kleiber, G., & Vuillaume, M. (2011). Sémantique des odeurs. *Langages*, 181, 17–37. DOI: <https://doi.org/10.3917/lang.181.0017>
- Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain and Language*, 81, 205–223. DOI: <https://doi.org/10.1006/brln.2001.2518>
- Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics*, 20(1), 1–24. DOI: <https://doi.org/10.1016/j.jneuroling.2006.02.001>



- Klepousniotou, E., Pike, G. B., Steinhauer, K., & Gracco, V. (2012). Not all ambiguous words are created equal: An EEG investigation of homonymy and polysemy. *Brain and Language*, 123(1), 11–21. DOI: <https://doi.org/10.1016/j.bandl.2012.06.007>
- Klepousniotou, E., Titone, D., & Romero, C. (2008). Making sense of word senses: The comprehension of polysemy depends on sense overlap. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1534–1543. DOI: <https://doi.org/10.1037/a0013012>
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Lindsay, M., & Aronoff, M. (2013). Natural selection in self-organizing morphological systems. In N. Hathout, F. Montermini & J. Tseng (Eds.), *Morphology in Toulouse: Selected Proceedings of Décembrettes 7* (pp. 537–556). Lincom Europa.
- Lombard, A., Barque, L., Huyghe, R., & Gras, D. (2023). Regular polysemy and novel word-sense identification. *The Mental Lexicon*, 18(1), 94–119. DOI: <https://doi.org/10.1075/ml.21002.lom>
- Lopukhina, A., Laurinavichyute, A., Lopukhin, K., & Dragoy, O. (2018). The mental representation of polysemy across word classes. *Frontiers in Psychology*, 9. DOI: <https://doi.org/10.3389/fpsyg.2018.00192>
- Miller, G. A. (2005). WordNet: A lexical database for English. *Communications of the ACM*, 38(11), 39–41. DOI: <https://doi.org/10.1145/219717.219748>
- Nunberg, G. (1995). Transfers of meaning. *Journal of Semantics*, 12(2), 109–132. DOI: <https://doi.org/10.1093/jos/12.2.109>
- Nunberg, G., & Zaenen, A. (1992). Systematic polysemy in lexicology and lexicography. In H. Tommola, K. Varantola, T. Salmi-Tolonen & J. Schopp (Eds.), *Proceedings of the Euralex II* (pp. 386–396). University of Tampere.
- Pinker, S. (1999). *Words and rules: The ingredients of language*. Weidenfeld and Nicolson.
- Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28, 73–193. DOI: [https://doi.org/10.1016/0010-0277\(88\)90032-7](https://doi.org/10.1016/0010-0277(88)90032-7)
- Plag, I. (1999). *Morphological productivity: Structural constraints in English derivation*. Mouton de Gruyter.
- Pustejovsky, J. (1995). *The Generative Lexicon*. MIT Press.
- Pylkkänen, L., Llinás, R., & Murphy, G. L. (2006). The representation of polysemy: MEG evidence. *Journal of Cognitive Neuroscience*, 18(1), 97–109. DOI: <https://doi.org/10.1162/089892906775250003>
- Rabagliati, H., & Snedeker, J. (2013). The truth about chickens and bats: Ambiguity avoidance distinguishes types of polysemy. *Psychological Science*, 24(7), 1354–1360. DOI: <https://doi.org/10.1177/0956797612472205>
- Renouf, A. (2013). A finer definition of neology in English: The life-cycle of a word. *Linguistics*, 57, 177–208. DOI: <https://doi.org/10.1075/scl.57.14ren>

- Rodd, J., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, G., & Davis, M. H. (2012). Learning new meanings for old words: Effects of semantic relatedness. *Memory & Cognition*, *40*, 1095–1108. DOI: <https://doi.org/10.3758/s13421-012-0209-1>
- Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*(2), 245–266. DOI: <https://doi.org/10.1006/jmla.2001.2810>
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464. DOI: <https://doi.org/10.1214/aos/1176344136>
- Smyk-Bhattacharjee, D. (2009). *Lexical innovation on the internet. Neologisms in blogs* [dissertation]. University of Zurich.
- Spencer, A. (2019). The nature of productivity (including word formation versus creative coining). *Oxford research encyclopedia of linguistics*. DOI: <https://doi.org/10.1093/acrefore/9780199384655.013.587>
- Srinivasan, M., & Rabagliati, H. (2015). How concepts and conventions structure the lexicon: Cross-linguistic evidence from polysemy. *Lingua*, *157*, 727–779. DOI: <https://doi.org/10.1016/j.lingua.2014.12.004>
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*(6), 645–659. DOI: [https://doi.org/10.1016/S0022-5371\(79\)90355-4](https://doi.org/10.1016/S0022-5371(79)90355-4)
- Twilley, L. C., Dixon, P., Taylor, D., & Clark, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. *Memory & Cognition*, *22*(1), 111–126. DOI: <https://doi.org/10.3758/BF03202766>
- Van de Velde, D. (1995). *Le spectre nominal: des noms de matières aux noms d'abstractions*. Peeters.
- Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2005). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. <https://journals.sagepub.com/doi/10.1080/17470218.2013.850521>. DOI: <https://doi.org/10.1080/17470218.2013.850521>
- Wittgenstein, L. (1958). *Preliminary studies for the "Philosophical Investigations", generally known as the blue and brown books*. Blackwell Publishers Ltd.
- Yurchenko, A., Lopukhina, A., & Dragoy, O. (2020). Metaphor is between metonymy and homonymy: Evidence from event-related potentials. *Frontiers in Psychology*, *11*. DOI: <https://doi.org/10.3389/fpsyg.2020.02113>

