



CAOSS and transcendence: Modeling role-dependent constituent meanings in compounds

Fritz Günther¹  · Marco Marelli^{2,3}

Received: 7 November 2020 / Accepted: 23 August 2021
© The Author(s) 2021

Abstract

Many theories on the role of semantics in morphological representation and processing focus on the interplay between the lexicalized meaning of the complex word on the one hand, and the individual constituent meanings on the other hand. However, the constituent meaning representations at play do not necessarily correspond to the free-word meanings of the constituents: Role-dependent constituent meanings can be subject to sometimes substantial semantic shift from their corresponding free-word meanings (such as *-bill* in *hornbill* and *razorbill*, or *step-* in *stepmother* and *stepson*). While this phenomenon is extremely difficult to operationalize using the standard psycholinguistic toolkit, we demonstrate how these as-constituent meanings can be represented in a quantitative manner using a data-driven computational model. After a qualitative exploration, we validate the model against a large database of human ratings of the meaning retention of constituents in compounds. With this model at hand, we then proceed to investigate the internal semantic structure of compounds, focussing on differences in semantic shift and semantic transparency between the two constituents.

Keywords Compound words · Psycholinguistics · Compound constituents · Semantic shift · Compound representation

Although finding an exact, commonly shared definition of *compound words* remains a theoretical challenge (Bauer, 1998; Lieber & Štekauer, 2009; Plag, 2006), a central defining feature is that they are combinations of (at least) two constituents that can themselves be used as free-standing words (i.e., the compound *foxhound* contains the constituents *fox* and *hound*). Within the compound, these constituents take on specific roles: The head (typically the second constituent in Germanic languages such

✉ F. Günther

¹ University of Tübingen, Tübingen, Germany

² University of Milano–Bicocca, Milan, Italy

³ NeuroMI, Milan Center for Neuroscience, Milan, Italy

as English; Williams, 1981) defines the syntactic and semantic category of the compound (a *foxhound* is a type of *hound*), and the modifier further specifies the head meaning (a *foxhound* can be used for hunting foxes). This inherent internal structure renders compounds highly interesting objects of research in both linguistics and psychology of language (Gagné, 2009; Libben, 2006, 2017; Lieber & Štekauer, 2009). In the cognitively-oriented literature on compound representation and processing, many studies have taken a special interest in the interplay between the meanings of the constituents and the compound: Traditionally such semantic influences were a crucial control variable when investigating purely morphological effects in processing, but semantic effects also increasingly became the focus of investigation themselves (for overviews, see Amenta & Crepaldi, 2012; Günther & Marelli, 2019; Schäfer, 2018; Schmidtke et al., 2018).

The standard approach in this context is to define central theoretical concepts such as semantic transparency (the degree to which a compound meaning can be related to its constituent meanings; Sandra, 1990) or compositionality (the degree to which a compound meaning can be predicted from its constituent meanings; Marelli & Luzzatti, 2012) by taking these constituent meanings “as is”: The implicit assumption is that the *fox* and the *hound* in *foxhound* are the same *fox* and *hound* that we would encounter in sentences such as *The fox is a shy animal* or *The hound was trained for hunting*. This assumption is illustrated by the fact that judgments of meaning relatedness between a constituent and the compound meaning (as collected by for example Juhasz et al., 2015 or Kim et al., 2019) are the most common measure for semantic transparency; and it is particularly explicit when semantic transparency is measured as the similarity between meaning representations in the form of distributional vectors (Günther & Marelli, 2019; Schmidtke et al., 2018; Pham & Baayen, 2013), which are obtained directly from the distributions of the individual words (*fox*, *hound*, *foxhound*) in large text corpora (Landauer & Dumais, 1997; Mander et al., 2017).

Surely, in the majority of cases, compound meanings are at least partially informed by the constituent free-word meanings. Indeed, for the purpose of communication, this informativity is a rather useful property for a complex linguistic construction (Costello & Keane, 2000) – after all, there are good pragmatic reasons to call a hound used for hunting foxes *foxhound* instead of *olivegarden*. However, this does not necessarily imply that the constituents embedded in a compound, which give rise to its meaning, can simply be equated with their free-word counterparts (Bell & Schäfer, 2013; Libben, 2014; Schäfer & Bell, 2020). As argued by Libben (2017), word meanings are highly dependent on the specific context they are used in (see also Janssen, 2001; Kintsch, 2000), which is especially relevant when they are used as compound constituents.

Most compounds are part of morphological families that have a strong tendency to share semantic aspects. As an example, take the word *bill* in *shoebill*, *hornbill*, or *razorbill*. Although *bill* can be used synonymously to *beak*, this is rarely the intended meaning when *bill* is found as a free word – as a free word, it usually is used in a legislative sense (*The senate passed the bill*), in a financial sense (*I still have to pay this bill*), or sometimes as a first name (*Meet my friend Bill!*). However, when used as a compound head, the ornithological facet is dominant, and almost the entire morphological family of the head *bill* consists of birds. Similarly, the modifier *step*

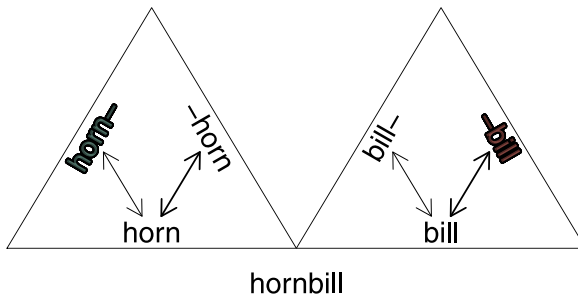


Fig. 1 Morphologically transcended representations of the words *horn* and *bill* (adapted from Libben, 2014). As-constituent representations are assumed to be linked through the free word meanings, with as-head representations being closer to the free word meanings than as-modifier representations. Representations highlighted in color are activated as constituents in the processing of the compound *hornbill* (Color figure online)

in *stepmom*, *stepson*, or *stepsister* has little to do with the act of walking, or with the steps on a stair. These examples also demonstrate that, although the meaning of constituents in a compound can be different from their original free-word meanings, this does not preclude a certain systematicity to these differences.

Observations like these lead to the formulation of the *morphological transcendence hypothesis* (Libben, 2014). According to this hypothesis, repeated experience with words in a constituent role can lead to the formation of morphologically transcended as-constituent meaning representations in the speaker's mind (such as the specific bird-related meaning of *-bill* when used as a head), which are connected but not identical to the free-word meaning representation (*bill*) (see Fig. 1). In other words, a compound such as *stepmother* would include as a constituent the morphologically transcended *step-* (shared with *stepson* or *stepdad*), rather than the *step* that is similar to *move* or *stand*. As evident in this example, the constituent meanings can experience an – at times dramatic – *semantic shift* in their role as modifiers or heads of compounds, and thus drift away from their original free-word meanings (see also Bell & Schäfer, 2013). According to Libben (2014), morphological transcendence has major theoretical implications, postulating that as-constituent representations rather than free-word representations should be considered as the psychologically relevant units for compound representation and processing.

However, high-level semantic concepts such as morphological transcendence or semantic shift are very elusive to the standard psycholinguistic toolkit. Consequently, studies subjecting these phenomena to empirical investigation are scarce. In the few existing studies (Libben et al., 2018; Smolka & Libben, 2017), word frequencies were used as a proxy measure, based on the assumption that there should be stronger competition between the free-word representation and the as-constituent representation for more frequent words. However, this measure is rather indirect, and relies on several theoretical assumptions that are only vaguely related to the concept in question (such as lexical competition and the interaction between competition and frequency).

The present study aims at closing the gap between the potential theoretical relevance of the morphological transcendence hypothesis (Libben, 2014) and the current state of empirical research, which can be attributed to the absence of a theoretical

conceptualization and methodological framework to adequately formalize the concepts in question. We demonstrate how role-specific as-constituent meanings can be directly represented in a quantitative format, by employing a computational model of compounding (Marelli et al., 2017) based on compositional distributional semantics (Baroni et al., 2014a). This allows us to obtain direct measures of semantic shift, which we initially explore in a qualitative analysis before validating them against participant judgments in the largest available database of English compounds (Gagné et al., 2019). After this evaluation, we then use these representations to empirically investigate several hypotheses by Libben (2014) on the internal semantic structure of compounds.

1 Modeling role-dependent constituent meanings

In this section, we describe our data-driven computational model to represent role-dependent constituent meanings. This model is implemented in a distributional semantics framework, and therefore based on substantial language experience, as approximated by large-scale corpora.

1.1 The word material

The word material forming the basis for our model and all empirical investigations in the present study consists of 4,429 closed-form compounds and their constituents. The compounds were collected from four dedicated closed-form compound databases in English: The 629-compound database by Juhasz et al. (2015), the 1,865-compound database by Günther and Marelli (2019), the 2,861-compound database by Kim et al. (2019), and the 8,376-compound database by Gagné et al. (2019). From this collection of compounds, we extracted those for which the compound and its constituents had a frequency higher than 50 in a large corpus (~ 2.8 billion words) of natural language (Baroni et al., 2014b): A concatenation of the web-collected ukWaC (Baroni et al., 2009), a 2008 English Wikipedia dump, and the print-media based British National Corpus (BNC Consortium, 2007). The frequency threshold was applied to guarantee that reliable model representations could be obtained for all words (i.e., compounds as well as their constituents; Li et al., 2019). Our semantic model will be trained on the exact same corpus (see the next section), which can serve as an approximation of speakers' language experience (Johns et al., 2016).

For all the words in the word material, we collected compound frequencies and constituent frequencies (i.e., the frequencies of modifiers and heads when used as free words) from the 2.8-billion word corpus. Furthermore, deriving all measures from the same corpus instead of assembling it from different sources ensures comparability between the different measures (Hollis, 2017). The same reasoning applies to our collection of compound words, which will serve as the training set for our computational model of compounding (Marelli et al., 2017, see below) and thus be treated as a proxy for a speaker's experience with compounds. Consequently, we measure modifier and head family sizes as the number of compound types sharing the constituent in question in the respective position using our collection of compounds as the reference list. Additionally, we collected as-modifier and as-head frequencies

(i.e., morphological family frequencies) for each constituent as the number of compound tokens sharing the constituent in question in the respective position. Due to their heavily right-skewed distributions, all frequency and family size measures were log-transformed.

1.2 The semantic model

In our model, word meanings are represented as high-dimensional numerical vectors reflecting their distributional patterns in language, based on the assumption that these distributional patterns reflect or even determine the meanings of words (Lenci, 2008). In practice, these *distributional vectors* are obtained from large collections of text data (which serve as a proxy for the language experience of speakers) using learning algorithms (Jones et al., 2015). In empirical as well as theoretical studies, distributional vectors have been established as a powerful model for cognitively-plausible semantic representations (Günther et al., 2019), also for morphologically complex words (Amenta et al., 2020). In addition, we consider high-dimensional distributional vectors to fulfill the criteria set out by Libben (2017) for psychologically adequate representations of word meanings, as “a means by which we can conceptualize words in the mind in a manner that is concrete enough to provide a scaffolding for knowledge advancement and also dissimilar enough from words on pages, or dictionaries on desks to enable us to maintain a safe distance from the fallacy of misplaced concreteness” (p. 54).

The specific model employed here was constructed using the *cbow* version of the *word2vec* architecture (Mikolov et al., 2013a,b), a neural network model with one hidden layer aimed at predicting a word given its context (the n words surrounding the target word). The distributional vector for any target word is then computed as the activation values of the hidden layer for this given word input (see Fig. 2 for a graphical representation of the model architecture). The high performance of the *cbow* model in a variety of semantic tasks has been established in several systematic model evaluation studies (Baroni et al., 2014b; Mandera et al., 2017; Pereira et al., 2016).

Specifically, we employed for the best-performing model identified by Baroni et al. (2014b), who systematically explored the parameter space for models in the *word2vec* architecture: 400-dimensional vectors trained to predict words within a 5-word window (negative sampling with $k = 10$, subsampling with $t = 1e^{-5}$). The model was trained on the same ~ 2.8 billion word corpus described in the *word material* section. As noted there, a frequency threshold of 50 was employed to guarantee reliable vector representations for all words (see Li et al., 2019).

Note that distributional models do *not* represent meanings as single, clear-cut symbolic units, but instead in distributional terms (which is especially important for the subsequently discussed compositional and as-constituent meanings). Having exactly one vector representation for each word does not imply monosemy, and many different meaning aspects can be encoded in a single distributional representation (for more detailed discussions, see Günther et al., 2019; Kintsch, 2007). In fact, the distributional representation format does not commit to any notion of discrete meaning representations or interpretations (such as, for example, closed sets of WordNet senses; Miller, 1995). Instead, a distributional vector indicates a point in a semantic space, which can more or less close to several different meanings. The vector for a

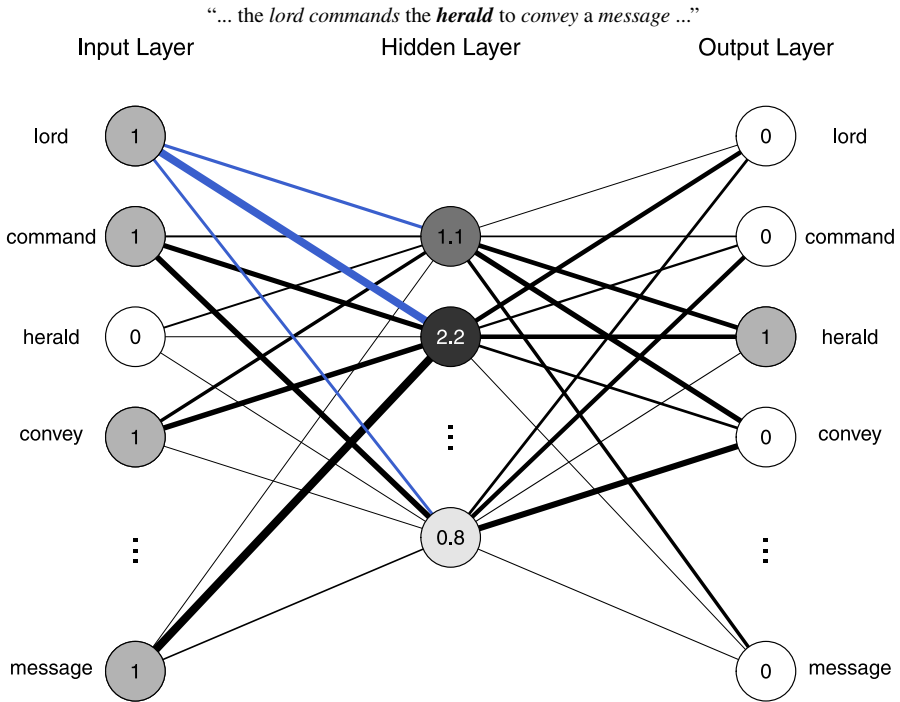


Fig. 2 A snapshot of the *cbow* model, in the training step of processing the target word *herald* in the utterance “... the lord commands the *herald* to convey a message ...” (function words ignored for this example). The widths of the edges represent the current weights in the network. The vector representation for each word is taken as the activation of the hidden layer if only this word is active (set to 1) in the input, which exactly corresponds to the weights from this input nodes to the nodes in hidden layer (see the weights for *lord* in blue) (Color figure online)

word that is traditionally considered polysemous then defines a point that lies in between different “semantic neighborhoods”: For example, the vector for *mouse* lies in between a cluster of words related to computers, and a cluster of words related to animals (see Fig. 3 in Günther et al., 2019), although without explicitly conceptualizing these clusters as different word senses or interpretations. A model that can link such distributional representations to discrete word senses was recently proposed by Rodd (2020), in which a form-based blend state of different meanings (such as a distributional vector) acts as the entry point for meaning access that is subsequently attracted towards familiar, discrete interpretations by context information.

Constituent words (in their free-word versions) and compounds (in their whole-word meaning) are used as free words in natural text. Consequently, the *cbow* model allows us to represent both free-word constituent meanings and whole-word compound meanings as distributional vectors. However, since as-constituent meanings are by definition not used as free words, we need to extend our model in order to obtain representations for these meanings. To this end, we first introduce a compositional model (Guevara, 2010; Marelli et al., 2017), and then demonstrate how this model can be employed to achieve this goal.

1.3 Role-dependent constituent meanings in a compositional model

Distributional models, as described in the previous section, allow us to represent word meanings as attested in natural language contexts. However, compounding is an inherently productive phenomenon: Speakers of English can be expected to encounter novel compounds on an almost daily basis, and are able to understand their meaning “on the fly”, without apparent effort (Downing, 1977; Wisniewski, 1997). This requires a powerful meaning-combination process that is able to produce a compositional compound meaning from its constituent parts.

Such compositional compound meanings are (essentially by definition) not directly observable in natural language – novel compounds are defined as not being part of the previous language experience. In order to represent these compositional meanings, we adapt the CAOSS model (Compounding as Abstract Operation in Semantic Space, Marelli et al., 2017; see Figure 3 for a graphical illustration), which has recently been successfully employed to represent the compositional meaning of both novel (Günther & Marelli, 2020; Marelli et al., 2017) and familiar compounds (Günther & Marelli, 2019; Günther et al., 2020).¹ The CAOSS model computes the compositional meaning of compounds c from their constituent meanings u and v as

$$c = M \cdot u + H \cdot v, \quad (1)$$

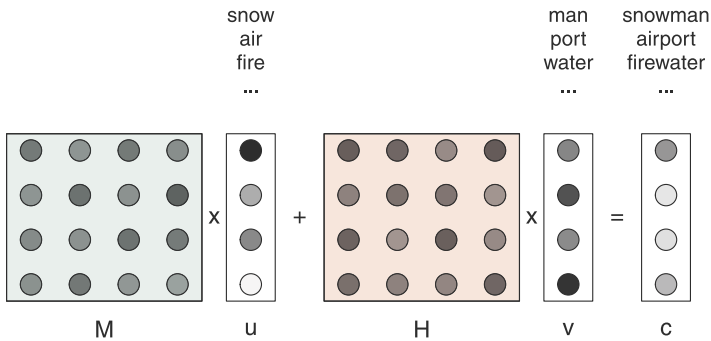
with M and H are $k \times k$ -dimensional weight matrices applied to the constituent vectors before they are added together (see also Guevara, 2010).

The weight matrices M and H are estimated in a training procedure (using the DISSECT toolkit; Dinu et al., 2013), with the 4,429 words in the word material described above serving as the training set. The model is only trained on closed-form compounds, because the LADEC database (Gagné et al., 2019) – against which it will be empirically evaluated – only contains closed-form compounds, and we wanted to avoid systematic differences between this item set and the training items. Furthermore, with this restriction on closed-form compounds our present work remains consistent with previous research on the CAOSS model (e.g. Günther & Marelli, 2019, 2020; Marelli et al., 2017). This is however not an inherent restriction of CAOSS, which could also be trained on open-form compounds in order to, for example, investigate phenomena related to spelling alterations (Kuperman & Bertram, 2013; Marelli et al., 2015).

Through its training procedure, the CAOSS model learns to extract a common, general-level meaning-combination structure from all the compounds it knows (that is, all the compounds in its training set; Marelli et al., 2017). The matrices M and H are estimated so that, on average, the whole-word compound vectors for the compounds in the training set are best predicted (applying a least-squares regression)

¹As an intuitive example for familiar compounds, it helps to imagine being a child or a foreign speaker of English. In this case, a reasonable interpretation for *ladybird* (the British English word for *ladybug*) is that it describes a kind of bird, maybe a female one – which is different from its actual meaning of being a little bug with dots.

CAOSS: Training



CAOSS: Composition

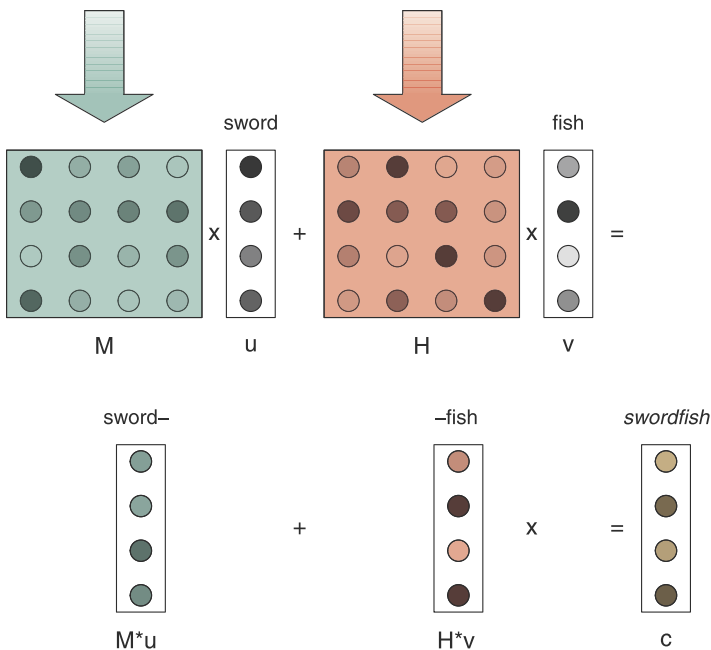


Fig. 3 Graphical illustration of the CAOSS model. The weight matrices are trained on a large training set of compounds, and then used to derive compositional vector representations for any combination of constituents. As illustrated in the last row, these weight matrices are used to update constituent vectors into role-dependent as-constituent vectors before combining them

from their constituent vectors u and v , following Equation (1). As a simple toy example, assume that there are four constituent vectors $a = (1, 1)$, $b = (2, 4)$, $c = (1, 0)$ and $d = (0, 2)$, as well as two compound vectors $ab = (7, 20)$ and $cd = (2, 8)$. In

that case, ab and cd can be perfectly predicted from their constituents by setting $M = \begin{pmatrix} 2 & 3 \\ 2 & 2 \end{pmatrix}$ and $H = \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}$,² which would thus be the training result.

As can be seen in Equation (1), the CAOSS model implements compounding as a two-step process. The constituent meanings u and v are updated into their role-specific modifier and head meanings $M \cdot u$ and $H \cdot v$, respectively,³ before these two newly-obtained role-specific meanings are combined into a single meaning. Thus, role-dependent as-constituent representations are a central component of the CAOSS model, which therefore offers a unique opportunity to quantify these meanings and hence directly investigate them in empirical studies.

These model representations allow us to measure the semantic shift of constituent meanings, a central concept of Libben's (2014) morphological transcendence hypothesis. More specifically, we measure *modifier consistency* and *head consistency* as the cosine similarity between a constituent's free-word vector (u and v , respectively) and its as-constituent vector ($M \cdot u$ and $H \cdot v$, respectively) for the 1,128 different words used as modifiers and 883 different words used as head in our word material. For the purpose of clarity, we adopt the term *consistency* instead of *semantic shift*, since the cosine is a measure of similarity between two meanings, while *shift* implies dissimilarity. Hence, *semantic shift* – the change of a word meaning when used in a specific constituent role – would be defined as $1 - \text{consistency}$.

2 Model evaluation

Our model, being able to compute some positionally-updated constituent representations during the compositional process does not, by itself, guarantee that these can be taken for as-constituent representations at face value. Focusing on the most salient property of these representations – the possibility to measure semantic shift – we first provide qualitative examples to examine their intuitive plausibility, before following up with a more systematic quantitative analysis. Data and analysis scripts are available at the OSF repository for this project (<https://osf.io/jxyrn>).

2.1 Qualitative examination

To provide an intuitive basis to examine semantic shift as indicated by our model, ten constituents scoring extremely high or low on semantic consistency (within the top or bottom 10%) are provided in Table 1. These examples are representative of large parts of the item set, which can be explored in its entirety at the OSF repository for this project.

As can be seen in Table 1, the model predictions are overwhelmingly in line with intuitions about semantic shift: For example, the *worm-* in the two members of its morphological family – *wormwood* (a plant used to produce absinth) and *wormhole*

²For ab , we get $M \cdot a + H \cdot b = (2 \cdot 1 + 3 \cdot 1, 2 \cdot 1 + 2 \cdot 1) + (1 \cdot 2 + 0 \cdot 4, 2 \cdot 2 + 3 \cdot 4) = (5, 4) + (2, 16) = (7, 20)$, and analogously for cd .

³In the previous example, $a = (1, 1)$ is updated into $M \cdot a = (5, 4)$, and $b = (2, 4)$ is updated into $H \cdot b = (2, 16)$.

Table 1 Constituents with very low or very high consistency scores (all within the top or bottom 10% with respect to these values). Cases with only one example have a family size of one in our dataset

modifier consistency			
very low		very high	
nick-	(<i>nickname</i>)	skate-	(<i>skateboard, skateboarder</i>)
worm-	(<i>wormwood, wormhole</i>)	chairman-	(<i>chairmanship</i>)
tom-	(<i>tomboy, tomcat</i>)	stair-	(<i>staircase, stairwell</i>)
white-	(<i>whitewash, whitetail</i>)	beach-	(<i>beachside, beachwear</i>)
dove-	(<i>dovecote, dovetail</i>)	clergy-	(<i>clergyman</i>)
see-	(<i>seesaw</i>)	cavalry-	(<i>cavalryman</i>)
over-	(<i>overdrive, overflow</i>)	folk-	(<i>folksinger, folklore</i>)
bag-	(<i>bagpipe, bagman</i>)	farm-	(<i>farmhouse, farmland</i>)
red-	(<i>redcoat, redwood</i>)	news-	(<i>newsagency, newsfeed</i>)
nit-	(<i>nitpick</i>)	rail-	(<i>railway, railtour</i>)
head consistency			
very low		very high	
-back	(<i>paperback, throwback</i>)	-daughter	(<i>goddaughter, granddaughter</i>)
-court	(<i>ballcourt, crosscourt</i>)	-journalist	(<i>photojournalist</i>)
-cote	(<i>dovecote, bellcote</i>)	-writer	(<i>ghostwriter, scriptwriter</i>)
-out	(<i>lookout, dropout</i>)	-flour	(<i>cornflour</i>)
-press	(<i>wordpress, winepress</i>)	-fish	(<i>starfish, codfish</i>)
-round	(<i>background, newsround</i>)	-boat	(<i>lifeboat, motorboat</i>)
-crest	(<i>firecrest, goldcrest</i>)	-scorer	(<i>goalscorer, topscorer</i>)
-peter	(<i>salt peter</i>)	-mill	(<i>sawmill, windmill</i>)
-bill	(<i>shoebill, hornbill</i>)	-father	(<i>godfather, stepfather</i>)
-saw	(<i>jigsaw, seesaw</i>)	-flow	(<i>workflow, bloodflow</i>)

(a tunnel connecting points in timespace, or a part of a pinball table) – shows no obvious link to the small animal that is a *worm*.

However, this does not imply that the meaning of low-consistency constituents is completely idiosyncratic in every single compound of its family: The *-bill* in *shoebill* or *hornbill* is used almost exclusively to refer to birds (as in other examples such as *spoonbill*, *thornbill*, or *broadbill*) or at least animals with beaks (*duckbill*, describing a platypus), and so experiences a semantic shift that is substantial and systematic at the same time. Thus, while the *beak* meaning of *bill* is extremely underrepresented when used as a free word (here, the meanings related to either payment or legislation are dominant), it becomes dominant when used in compounds and can be used very productively as a *pars pro toto* to refer to birds. Exploring the neighborhoods of these words further illustrates this phenomenon: The five nearest words to *bill* are *bills*, *legislation*, *amendment*, *act*, and *reauthorization*. This changes dramatically for the as-constituent meaning *-bill*, for which the five nearest words are *-bird*, *-tail*, *-lark*, *-catcher* (as in *flycatcher* and *oystercatcher*) and *-cap* (as in *blackcap* and *redcap*). In addition to these bird-denoting compound heads, the 50 nearest neighbors further

include several bird names in their free-word forms, such as *woodswallow*, *aerodramus*, *ruficollis*, *haematopus*, and *gallinula*. This captures, on a clearly quantifiable level, the shift of the *bill* meaning from a legal to an ornithological connotation when the word is used as the head constituent in a compound.

On the other hand, other words almost completely maintain their original free-word meanings when entering into compounds. One prime example are family members: A *goddaughter* is a daughter “before god”, a *granddaughter* is a daughter of one’s child, and a *stepdaughter* is a daughter of one’s spouse (analogously for *-father*, *-brother*, or *-mother*). This can also be the case for modifiers: For example, all compounds including the modifier *stair-* are used to refer to stairs or something containing stairs (*staircase*, *stairwell*, *stairway*). This is again reflected in the neighborhoods of these words: *stair-* is the fourth-nearest neighbor to *stair*, after *staircase*, *stairs*, and *staircases*. Accordingly, the neighborhood of *stair-* is very similar to that of *stair*, with *stair* itself being in the sixth position after *staircase*, *stairway*, *door-*, *entrance-*, and *gate-*.

These qualitative examples illustrate that the model’s predictions on the semantic shift of constituents are indeed plausible on an intuitive level, giving a first indication for the validity of the model-obtained as-constituent representations.

2.2 Quantitative analysis

To back up these qualitative intuitions with a systematic evaluation of our model’s capacity to capture semantic shift, we conducted a quantitative analysis in which we validate our measures of semantic consistency against participant ratings of constituent meaning retention.

2.2.1 Methods

We employed data from a large-scale study on compounds by Gagné et al. (2019). In this study, participants were presented with sentences such as “How much does *bed* retain its meaning in *flowerbed*?” or “How much does *flower* retain its meaning in *flowerbed*?”. Participants gave their answer by positioning a slider on a continuous scale with the endpoints “retains none of its meaning” (coded as 0) and “retains all of its meaning” (coded as 100). This instruction by Gagné et al. (2019) comes very close to intuitively describing the concept of semantic shift to naïve participants, since it emphasizes the relation between the original meaning of a constituent and its meaning within a given compound. Each item was rated by 21 to 44 participants, who each provided ratings for between 118 and 150 compounds.

Since semantic shift as defined by Libben (2014) is a constituent-specific rather than a compound-specific measure (see Fig. 1), we computed a constituent-level meaning retention score for all constituents in our dataset by averaging the meaning retention scores for all compounds in the respective morphological families (i.e., the score for *-bill* is computed by averaging over *hornbill*, *shoebill*, etc.). This results in average meaning retention ratings for a set of 1,141 different modifiers and 941 different heads.

Table 2 Parameters for the linear regression model predicting meaning retention ratings for modifiers and heads, as reported in the main analysis

modifier ratings			
parameter	<i>b</i>	<i>t</i>	<i>p</i>
intercept	13.00	3.50	< .001
modifier consistency	43.40	8.64	< .001
modifier frequency	2.27	9.00	< .001
modifier family size	-0.92	-1.23	.220
as-modifier frequency	1.15	3.21	.001
modifier length	1.99	5.04	< .001
head ratings			
parameter	<i>b</i>	<i>t</i>	<i>p</i>
intercept	21.38	5.56	< .001
head consistency	42.15	8.91	< .001
head frequency	2.54	9.68	< .001
head family size	0.51	0.69	.492
as-head frequency	-0.35	-1.01	.313
head length	2.45	6.02	< .001

2.2.2 Results

In an initial analysis, we observed medium correlations between the average meaning retention for modifiers and modifier consistency ($r = .31$, $t(1139) = 11.12$, $p < .001$), as well as between the average meaning retention for heads and head consistency ($r = .32$, $t(939) = 10.29$, $p < .001$).

To examine whether our model-derived consistency measures predict meaning retention over and above constituent-level lexical variables, we then estimated a linear regression model predicting modifier ratings from modifier consistency, free-word frequency, family size, and family frequency (as-constituent frequency), as well as length. The model results are displayed in the upper part of Table 2. As can be seen, there is a significant positive effect of modifier consistency in this model ($t = 8.64$, $p < .001$; see also the left panel of Fig. 4). Further, the model including the modifier consistency parameter explains the data significantly better than a model without this parameter, as indicated by a likelihood-ratio test ($F(1, 1135) = 74.56$, $p < .001$, with the explained variance increasing from $R^2 = .157$ to $R^2 = .209$). In an analogous model predicting head ratings from the head-related measures, we find a positive effect of head consistency ($t = 8.91$, $p < .001$, see the lower part of Table 2 and the right panel of Fig. 4), and the model including the head consistency parameter outperforms a model without it ($F(1, 935) = 79.41$, $p < .001$, with the explained variance increasing from $R^2 = .144$ to $R^2 = .211$).

To ensure that these are genuine role-specific effects and not some methodological or computational artifact, we performed a follow-up analysis to investigate whether meaning retention ratings can also be predicted from the “wrong” consistency score (that is, if modifier consistency predicts head meaning retention ratings, and vice versa). If our measures are indeed role-specific, this should not be the case. To this

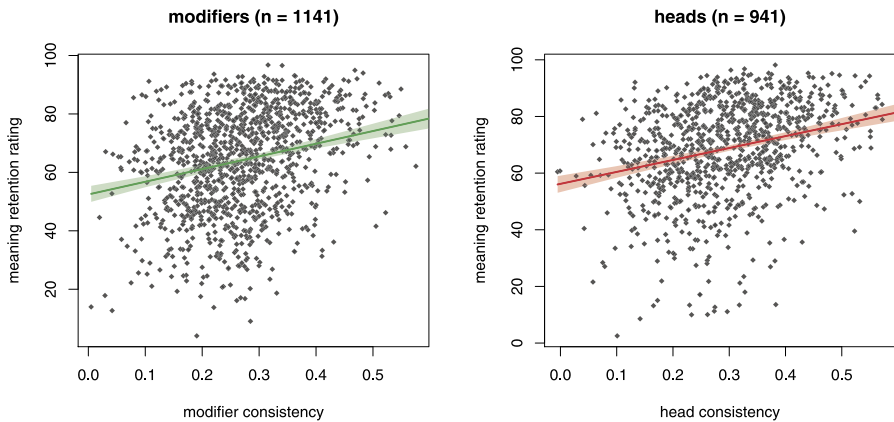


Fig. 4 Effect of the consistency effects on meaning retention ratings (left panel: modifier consistency and modifier ratings; right panel: head consistency and head ratings), with their respective .95-Wald confidence intervals

end, we applied the wrong weight matrix to the respective constituents, and then computed the similarities $\cos(u, H \cdot u)$ for every modifier u and $\cos(v, M \cdot v)$ for every head v .

For the modifier meaning retention ratings, we estimated a linear regression model predicting the rating from a word's *head* consistency as well as its modifier free-word frequency, modifier family size and as-modifier frequency, as well as its length. In this model, the effect of head consistency – that is, the incorrect consistency score – was significant ($t = 2.53$, $p = .012$). However, in this case the model fit ($R^2 = .162$) was far lower than in the model containing modifier consistency instead of head consistency as a predictor, which was described in the main analysis ($R^2 = .209$). Furthermore, in a model that contained both modifier and head consistency as predictors for modifier meaning retention ratings, in addition to the modifier-related lexical variables, only the effect of modifier consistency was significant ($t = 8.29$, $p < .001$ for modifier consistency; $t = -0.97$, $p = .332$ for head consistency).

We performed the analogous analysis for the head meaning retention ratings. In the model containing modifier consistency as well as all head-related lexical variables as predictors, the effect of modifier consistency was not significant in the first place ($t = 1.19$, $p = .236$). Naturally, the explained variance of this model ($R^2 = .145$) was considerably lower than for the model in the main analysis, in which the head consistency effect was significant ($R^2 = .211$). In the model that contained both modifier and head consistency as well as all head-related lexical variables, the effect of modifier consistency was significant but negative, while the much stronger effect of head consistency effect was positive ($t = 9.15$, $p < .001$ for head consistency; $t = -2.37$, $p = .018$ for modifier consistency). Since the modifier and head consistency scores for the 941 different words used as heads are moderately correlated ($r = .38$, $t(939) = 12.68$, $p < .001$), this transition from a non-significant positive parameter value to a significant negative value can be attributed to a statistical suppression effect.

Taken at face value, this correlation between our model-derived modifier and head consistency measures described in the previous analysis could be interpreted as our model being unable to correctly differentiate between the modifier and head role. To investigate this possibility, we further examined the 427 words of the LADEC dataset (Gagné et al., 2019) that are used in both roles (modifier and head) in our dataset. In this subset of the data, the correlation between modifier and head consistency ($r = .38$, $t(425) = 8.41$, $p < .001$) is very similar to the one reported for the 941 different heads. However, in the same dataset of 427 words, we also observe a very similar significantly positive correlation when comparing the modifier and head meaning retention ratings for the same words ($r = .30$, $t(425) = 6.56$, $p < .001$): As indicated by a comparison of two nonoverlapping correlations based on dependent groups as implemented in the R package *cocor* (Diedenhofen & Musch, 2015), there is no significant difference between the model-derived and the participant-based correlation ($z = 1.30$, $p = .193$). This demonstrates that our computational model does not produce an artificial relation between two unrelated measures; rather, it captures a systematic relation between modifier and head consistency that fleshes out human intuitions about these words, even though the model architecture was not explicitly designed to do so.

2.2.3 Discussion

This quantitative analysis provides evidence that the CAOSS architecture (Marelli et al., 2017) can indeed capture the semantic shift experienced by compound constituents, while specifically distinguishing between the two different constituent roles. In fact, our model-derived measures emerge as among the strongest predictors of meaning retention ratings (second only to frequencies, but far more impactful than other plausible candidates such as family size or family frequency), and are associated to a considerable increase in the explained variance. This implies that the weight matrices M and H capture this role-specific information as an emergent phenomenon of language experience.

An interesting side-observation of our study is the medium-level correlation between the semantic shift as a modifier and a head when a word is used in both constituent roles. This correlation in our model-derived measures is reflected in the participant ratings, which qualifies it as a genuine empirical phenomenon rather than a modeling artifact. This indicates that some words tend to be semantically more volatile across roles when used in a compound, and not only in one specific constituent role (extreme examples include *bill* (as in *billboard*, *hornbill*) or *pick* (as in *pickpocket*, *toothpick*), while others tend to be more semantically stable (extreme examples include *boat* (as in *boatyard*, *lifeboat*) or *gun* (as in *gunfire*, *handgun*)). Since we had no hypothesis with respect to this phenomenon, it is up to future work to identify relevant factors associated to this semantic volatility; possible candidates could be the diachronic origin of a word within the language, its diachronic usage as a constituent in compounds, the diachronic development of the various compounds it is used in, or its productivity as a constituent.

3 Investigations of semantic shift

Having established the validity of our model-derived measures of semantic shift (and, by extension, the as-constituent representations), we now employ our model to empirically investigate several hypotheses formulated by Libben (2014) concerning the internal semantic structure of compounds and their constituents.

3.1 Semantic shift by constituent role

One core assumption of the morphological transcendence hypothesis is that *different* as-constituent representations are formed depending on the constituent role in which a word is used (Libben, 2014; see Fig. 1 of the present article). That is, a given word will develop semantic peculiarities that will be different when used as the modifier constituent vis-à-vis as the head constituent of a compound (compare *bill-* as in *billboard* and *billhook* with *-bill* as in *hornbill* and *shoebill*). However, according to the morphological transcendence hypothesis, these different representations are still inherently connected to the representation of the same word, and thus remain linked through the free word meaning even though they may drift away from it in different directions. As an example, Libben (2014) discusses the case of *key*: In the modifier role, *key-* is often used in a metaphorical way (as in *keystone* or *keynote*). This use is afforded by the semantics of the free-word meaning (“*planning is key.*”). However, *-key* when used as a head does not take up on this meaning aspect (*turnkey*, *passkey*).

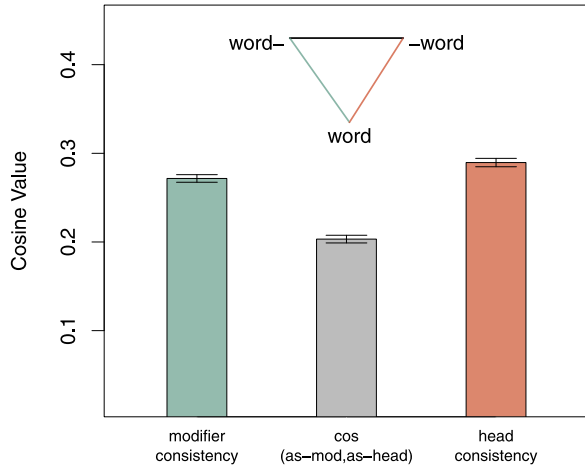
3.1.1 Drifting from the free-word meaning

If the free-word meaning is the starting point for both as-constituent meanings, and if the two as-constituent meanings do not influence one another in their drift from this free-word meaning (Libben, 2014), we expect the following pattern in terms of our model: For words which can be used both as a modifier and a head, both (a) the similarity between their as-modifier meaning (e.g., *bill-*) and free word meaning (*bill*) and (b) the similarity between their as-head meaning (*-bill*) and free word meaning (*bill*) are expected to be larger than (c) the similarity between their as-modifier meaning (*bill-*) and as-head meaning (*-bill*). In geometrical terms, the free word vector should lie between the as-modifier vector and the as-head vector, which are free to drift away from it in different directions. Note that this is an empirical hypothesis rather than a mathematical necessity: In principle, it could also be the case that both as-constituent meanings tend to drift away from the free-word meaning in the same direction. This would be reflected in very similar (or, in the most extreme case, identical) weight matrices M and H . If this was the case, we would observe the reverse pattern as the one predicted here: (c) would be larger than both (a) and (b).

To test the first hypothesis, we selected the 509 words that appeared both as modifier and head in our word material. For these words, the modifier consistency was, on average, higher than the similarity between the as-modifier meaning as the as-head meaning ($t(508) = 16.91$, $p < .001$). The same was true for head consistency ($t(508) = 20.21$, $p < .001$), see Fig. 5.⁴ These results are consistent with the hy-

⁴This pattern remains the same when we restrict the item set to “more established” constituent words with larger modifier and head family sizes than 1 (specifically, we tested family sizes of 2–5).

Fig. 5 Similarities between the as-constituent meanings and free word meanings, as measured by the cosine similarities between the respective distributional vector representations (means and standard errors)



pothesis that the as-constituent representations are linked through their free word meaning, and that both as-constituent meanings are more similar to this free-word meaning than they are to each other.

This pattern of results is in line with the assumption by Libben (2014) that both as-constituent meanings remain connected to the original free-word meaning, without assuming a direct connection between the two as-constituent meanings (see Fig. 1). That is, the direction of a word’s semantic shift when used a modifier does not necessarily influence the direction of its semantic shift as a head; and indeed, on average the as-constituent meanings drift away from each other. This implies that both as-constituent meanings can form independently from one another, each free to shift away from the free-word meaning in its own direction without being influenced by the other. Therefore, these results support the “trinity” of meaning postulated by the morphological transcendence hypothesis (see Fig. 1).

However, it should be noted again that while the *direction* of semantic shift in one constituent role is not determined by the other, the *amount* of semantic shift is: As reported in the quantitative analysis on semantic shift, there is a strong positive correlation between modifier and head consistency. This correlation is also found in the 509 words examined here ($r = .38$, $p < .001$). As argued above, this indicates that words which tend to retain much of their free-word meaning in the modifier role also retain much of it in the head role (such as *boat* or *wood*), and that words which experience substantial semantic shift in one role also do so in the other (such as *bill* or *face*).

3.1.2 Stronger shift for modifiers

A second claim put forward by Libben (2014) is the hypothesis that semantic shift is, on average, stronger for as-modifier meanings than for as-head meanings, and that “compound heads are linked in a more facilitatory manner to their corresponding whole word representations than modifier representations” (Libben, 2014, p. 20). Since the head of a compound typically determines the semantic category of a compound (a *songbird* is a kind of *bird*), the as-head meaning of *-bird* is expected to be

relatively close to the free word meaning of *bird*, while the specific function of the modifier can be more variable and sometimes less clear (see also Gagné & Shoben, 1997; Murphy, 1990).

In terms of our model, we expect that the head consistency is, on average, higher than the modifier consistency, for words which are used in both roles. We tested this on the same set of 509 words as described in the previous analysis. Indeed, the results of the analysis confirm the hypothesis ($t(508) = 3.55, p < .001$; see Fig. 5).

These results demonstrate that as-modifier meanings are subject to a higher semantic shift than as-head meanings, which retain more of their original free-word meaning. This pattern of results can easily be explained by very early models of conceptual combination: The Concept Spezialization Model (Murphy, 1988, 1990) views concepts as high-dimensional semantic feature vectors (distributional vectors can be seen as an abstraction of these representations; Günther & Marelli, 2016). In conceptual combination, the modifier then selectively affects specific dimensional values of the head. This will result in a majority of cases where the original head meaning is largely retained after combination (if not too many values are modified, or if the modification is not too large), while the modifier meaning will not be retained to the same degree. That is, the modifier is subject to more semantic shift than the head. This difference between constituent roles is not explicitly implemented in the CAOSS model – Equation (1) allows for, but does not necessitate different degrees of modulation of the constituent meanings by the weight matrices. However, the present results demonstrate that the model captures this property in a bottom-up, data-driven fashion, as an emergent feature of how the distributed architecture learns from natural language data.

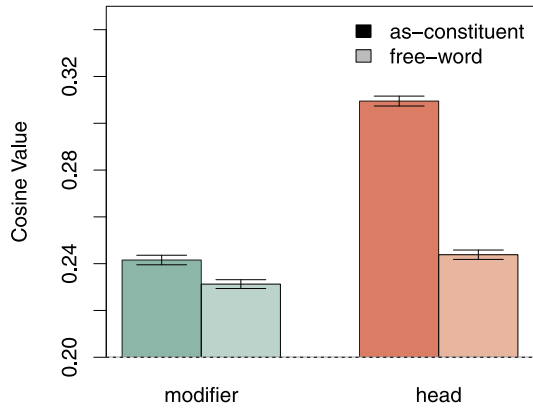
3.2 Semantic transparency by constituent role

In the previous section, we have obtained supporting evidence for the hypothesis that, on average, the head of a compound experiences a lesser degree of semantic shift than the modifier. Following up on the assumption that the meaning of compounds is derived from the as-constituent representations rather than the free-word meanings (Libben, 2010, 2014), Libben further hypothesizes that there should be fewer semantically opaque compound constituents with respect to the head position as compared to the modifier position.

This original hypothesis relies on a binary notion of semantic transparency (transparent vs. opaque), whereas our model measures semantic relations as graded, continuous variables. Since any cutoff value to artificially dichotomizes these graded values would be arbitrary, we re-formulate the hypothesis as follows: Compounds for which the head is more transparent than the modifier are expected to be more frequent than compounds for which the reverse is true, and head transparency should be higher on average than modifier transparency. If indeed as-constituent representations rather than free-word representations are the representational basis for compounds (Libben, 2014), then these differences should be more prominent when considering the as-constituent meaning variant.

In order to test these hypotheses, we initially measured the transparency of the head in terms of traditional semantic relatedness (Günther & Marelli, 2019;

Fig. 6 Constituent-wise semantic transparency in terms of relatedness: Similarities between the constituent meanings (both free-word and as-constituent representation) and whole-word compound meanings (means and standard errors)



Schmidtke et al., 2018): modifier relatedness was operationalized as the cosine similarity between the distributional vectors for the free-word modifier meaning (e.g., *song*) and the whole-word compound meaning (e.g., *songbird*), and analogously for head relatedness (here, the cosine similarity between *bird* and *songbird*).

When applying this operationalization, there were more compounds for which the head was more transparent than the modifier ($P = .52$, $p = .001$), and head relatedness was on average larger than modifier relatedness ($t(4428) = 4.83$, $p < .001$; see Fig. 6).

In a next step, we measure head and modifier transparency based on the as-constituent representations rather than the free-word representations of the constituent meanings (i.e., the cosine similarity between *song-* and *songbird*, and between *-bird* and *songbird*). When applying this conceptualization, there were considerably more compounds ($n_h = 2,784$) for which the head was more transparent than the modifier than compounds for which the reverse was true ($n_m = 1,645$; $P = .63$, $p < .001$ in a binomial test). This pattern was far more pronounced when considering the transparency measures based on as-constituent meanings as compared to measures based on free-word meanings ($z = 11.85$, $p < .001$ for the difference in a mixed-effects model).

The same is true for the comparison of average modifier and head transparency: When based on as-constituent meanings, head transparency was considerably higher than modifier transparency ($t(4428) = 22.63$, $p < .001$, see Fig. 6). Again, this pattern was far more pronounced when considering as-constituent meanings as compared to free-word meanings ($t(4428) = 21.37$, $p < .001$ for the interaction between type of transparency (modifier vs. head) and type of meaning representation (free-word vs. as-constituent) in a mixed-effects model).

Our model simulations are in line with Libben's (2014) hypothesis concerning the differences in semantic transparency by constituent role. As can be seen, these semantic transparency effects are more prominent when considering the units which, according to Libben (2014), are the ones relevant to compounding: as-constituent meaning, rather than free-word meanings. In line with this interpretation, as can be seen in Fig. 6, the as-constituent representations are overall more similar to the compound meaning than their free-word counterparts ($t(8667) = 5.19$, $p < .001$ in the

mixed-model analysis), highlighting the closer semantic connection between these morphologically transcended meanings and the compound meaning (see also Marelli et al., 2017).

4 Discussion

In this article, we present a model to capture the elusive role-dependent meanings representations of compound constituents, which play the central role in the theory of morphological transcendence by Libben (2014). More specifically, we employ the CAOSS model (Marelli et al., 2017), a large-scale and data-driven computational model of compounding, which computes these role-dependent as-constituent representations as an inherent part of its model architecture. In an initial qualitative analysis, we show that these representations appear to be plausible on an intuitive level. Critically, these intuitions are backed up by a quantitative analysis based on the human-based data by Gagné et al. (2019), which demonstrates that our model-derived measures of semantic shift predict participant ratings of constituent-meaning retention in compounds. Taken together, the results indicate the CAOSS as-constituent representations as promising direct and quantifiable operationalizations of role-dependent constituent meanings.

We then use these representations to empirically investigate several hypotheses on morphological transcendence (Libben, 2014), specifically on semantic shift and semantic transparency by constituent role. The results of these studies are in line with the hypotheses. On the one hand, this provides further empirical support for the validity of our modeling approach. On the other hand, it demonstrates that precise operationalizations of theoretical concepts, as presented here, are a valuable asset for investigating theoretical questions.

4.1 A system-based perspective on as-constituent meanings

The original formulation of the morphological transcendence hypothesis by Libben (2014) starts from psychological considerations of how constituent meanings of compounds are represented in the speakers' minds. However, at this point it is not clear if these meanings are either stored as permanent entries in semantic memory, or computed in an on-line process from the corresponding constituent free-word meanings whenever they are relevant. Our present studies were not designed to draw definitive conclusions in this respect. Still, the fact that the model proposed here appears able to adequately capture as-constituent meanings opens a dynamic and system-based perspective on the status of such elements.

In the original version of the morphological transcendence hypothesis by Libben (2014), morphological transcendence is the result of repeated experience with a specific constituent. This can then lead to idiosyncratic semantic shift, depending on the compounds in which the specific constituent is used. In contrast, our approach drastically deviates from this assumption by moving from the specific to the general level: Following the CAOSS model (Marelli et al., 2017), as-constituent meanings are derived from their corresponding free-word meanings during conceptual combination, via a role-dependent updating mechanism that is captured by two weight matrices.

Critically, these weight matrices are the *same* for all constituent combinations, and they are learned from *all* available compounds. Since individual as-constituent meanings are obtained by multiplying free-word vectors with a weight matrix (M or H), they will result from how individual word meanings interact with this general-level updating mechanism. This interaction can give rise to extreme semantic shift (*-bill*) or little semantic shift (*-boat*; see also Table 1), depending on the characteristics of the elements involved. Thus, also apparently idiosyncratic examples of semantic shift are in fact the result of a largely regular and predictable updating mechanism, as captured in the CAOSS weight matrices (similar to the “islands of reliability” in semantic opacity for affixed words, see Marelli & Baroni, 2015).

As a consequence, morphological transcendence and semantic shift result from experience with the entire compounding system, rather than with specific instances. This approach clearly places the idea of semantic shift at the center, and sidelines the notion of morphological transcendence (i.e., the formation of positionally-bound representations): If as-constituent meanings can be obtained by simply applying a general updating mechanism to the corresponding free-word meanings, it stands in question why additional positionally-bound representations would need to be stored as individual units in semantic memory. This view thus renders the updating into as-constituent meanings an inherent part of a general-level meaning-combination process, rather than postulating the existence of separately stored as-constituent representations that have to be accessed during processing.

This also directly addresses open issues of the original version of the morphological transcendence hypothesis, which assumes that as-constituents representations would develop as a function of repeated experience in the form of morphological families: In this original version, it remains unclear why a separate representation should be formed for the large-family, highly consistent *-fish*, but not for the small-family, strongly shifted *nick-*. By directly focusing on semantic shift as an inherent property and result of a general meaning-combination system, without assuming separately stored, morphologically transcended representations, this issue is no longer relevant.

Note that, although this view stands in contrast to some of the aspects of the initial morphological transcendence hypothesis by Libben (2014), it is entirely in line with its core underlying argument that not storage efficiency, but the maximization of opportunity for comprehension is the relevant factor in compound representation and processing: There is no principled reason to assume that accessing an additional positionally-bound representation in memory (that might even compete for activation; see Baayen et al., 2011; Schmidtke et al., 2018) should be faster than accessing the free-word meaning and applying a simple and general-level updating rule. In addition, the second option is more flexible, since it handles new constituents in the very same way as familiar ones, while at the same time being more efficient, since only free-word meanings (and two general-level updating functions) need to be stored instead of potentially three different representations for each constituent.

4.2 Emergent phenomena in a data-driven model

Throughout the present article, we have emphasized the entirely data-driven nature of our model. This means that, while the general structure of the algorithms (i.e., the

model architecture of the *cbow* model which produces semantic representations, and Equation (1) for meaning combination) are fixed, all parameter values are specified during training on actual data (the *cbow* model is trained on a corpus of natural language, and the CAOSS model on all compound words within that corpus). In this context, the corpus at the basis of these training procedures serves as a proxy representing aggregated language experience (on a population level, Günther et al., 2019, not necessarily at the level of individual speakers, see Schmidtke et al., 2018). The algorithms then play the role of learning models that aim at extracting meaning, and derive semantic representations from this experience (Günther et al., 2019; Landauer & Dumais, 1997; Jones & Mewhort, 2007; Marelli et al., 2017).

A direct consequence of this setup is that all representations and phenomena related to these representations (such as the similarities between them) emerge naturally as a function of the input (compare Rumelhart et al., 1986). In the absence of hand-coding, the model predictions are not influenced by pre-conceptions of “how things ought to be”. This complete reliance on the training data could in principle result in cases where the model simply does not produce the phenomena that are central to our theoretical considerations: While our model allows for these phenomena to emerge, it does not force them to. For example, one could end up with a case where the weight matrices M and H are identical (i.e., $M = H$), indicating a symmetric compounding system where the meaning of compounds such as *houseboat* and *boathouse* would be the same (opposed to the view that compounds are inherently asymmetric, Di Sciullo, 2005). However, our analyses on semantic shift by constituent role indicate that this is not the case. Similarly, one could even end up with a scenario where the weight matrices M and H are identity matrices, and the combination procedure is equivalent to a simple vector addition $c = u + v$ (Mitchell & Lapata, 2010). In this case, the concept of semantic shift would be non-existent for the model, which is again contrary to our findings. Therefore, distinctions between concepts such as free-word meanings and role-specific constituent meanings are a reflection of the inherent nature of the (English) compounding system as observed in natural language use.

4.3 Conclusion

In the present article, we have proposed a fully implemented, computational system to represent role-dependent as-constituent meanings within compounds, from which we derive quantitative measures of semantic shift. Importantly, this system is neither hand- nor hard-coded, but instead dynamically learns all the relevant representations from experience in a data-driven, bottom-up manner. As illustrated by the empirical studies presented here, this computational approach allows us to directly test new hypotheses and thus advance our theoretical understanding of the semantics of compound words.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by a Research Fellowship (no. 392225719) from the German Research Foundation (DFG), awarded to Fritz Günther.

Data Availability Data and analysis scripts are available at <https://osf.io/jxym>.

Declarations

Conflict of Interest The authors declare that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: an analytical review of morphological effects in visual word identification. *Frontiers in Psychology*, 3, 232.
- Amenta, S., Günther, F., & Marelli, M. (2020). A (distributional) semantic perspective on the processing of morphologically complex words. *The Mental Lexicon*, 15, 62–78.
- Baayen, R. H., Milin, P., Filipović Đurđević, D., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118, 438–481.
- 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, 209–226.
- Baroni, M., Bernardi, R., & Zamparelli, R. (2014a). Frege in space: A program for compositional distributional semantics. *Linguistic Issues in Language Technologies*, 9(6), 5–110.
- Baroni, M., Dinu, G., & Kruszewski, G. (2014b). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of ACL 2014 (52nd annual meeting of the Association for Computational Linguistics)*, pp. 238–247. East Stroudsburg: ACL.
- Bauer, L. (1998). When is a sequence of noun + noun a compound in English? *English Language and Linguistics*, 2, 65–86.
- Bell, M. J., & Schäfer, M. (2013). Semantic transparency: challenges for distributional semantics. In *Proceedings of the IWCS 2013 workshop towards a formal distributional semantics* (pp. 1–10). ACL.
- BNC Consortium (2007). The British National Corpus, version 3. Oxford: Bodleian Libraries. <http://www.natcorp.ox.ac.uk/>.
- Costello, F. J., & Keane, M. T. (2000). Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science*, 24, 299–349.
- Diedenhofen, B., & Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS ONE*, 10(4), e0121945.
- Dinu, G., Pham, N., & Baroni, M. (2013). DISSECT: DIStributIonal SEMantics Composition Toolkit. In *Proceedings of the system demonstrations of ACL 2013 (51st annual meeting of the Association for Computational Linguistics)* (pp. 31–36). East Stroudsburg, PA: ACL.
- Di Sciullo, A. M. (2005). *Asymmetry in morphology*. Cambridge, MA: MIT Press.
- Downing, P. (1977). On the creation and use of English compound nouns. *Language*, 53, 810–842.
- Gagné, C. (2009). Psycholinguistic perspectives. In R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of compounding* (pp. 255–271). New York, NY: Oxford University Press.
- Gagné, C. L., & Shoben, E. J. (1997). Influence of thematic relations on the comprehension of modifier–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 71.
- Gagné, C. L., Spalding, T. L., & Schmidtke, D. (2019). LADEC: the large database of English compounds. *Behavior Research Methods*, 51, 2152–2179.
- Guevara, E. (2010). A regression model of adjective–noun compositionality in distributional semantics. In *Proceedings of the 2010 workshop on geometrical models of natural language semantics* (pp. 33–37).

- Günther, F., & Marelli, M. (2016). Understanding Karma police: The perceived plausibility of noun compounds as predicted by distributional models of semantic representation. *PLoS ONE*, 11(10). <https://doi.org/10.1371/journal.pone.0163200>.
- Günther, F., & Marelli, M. (2019). Enter Sandman: Compound processing and semantic transparency in a compositional perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45, 1872–1882.
- Günther, F., & Marelli, M. (2020). Trying to make it work: Compositional effects in the processing of compound “nonwords”. *Quarterly Journal of Experimental Psychology*, 73, 1082–1091.
- Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*, 14, 1006–1033.
- Günther, F., Petilli, M. A., & Marelli, M. (2020). Semantic transparency is not invisibility: A computational model of perceptually-grounded conceptual combination in word processing. *Journal of Memory and Language*, 112, 104104.
- Hollis, G. (2017). Estimating the average need of semantic knowledge from distributional semantic models. *Memory & Cognition*, 45, 1350–1370.
- Janssen, T. M. V. (2001). Frege, contextuality and compositionality. *Journal of Logic, Language and Information*, 10, 115–136.
- Johns, B. T., Jones, M., & Mewhort, D. J. (2016). Experience as a free parameter in the cognitive modeling of language. In *Proceedings of the 38th annual conference of the Cognitive Science Society* (pp. 2291–2296). Austin, TX: Cognitive Science Society.
- Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114, 1–37.
- Jones, M. N., Willits, J., & Dennis, S. (2015). Models of semantic memory. In J. Busemeyer, Z. Wang, J. Townsend, & A. Eidels (Eds.), *Oxford handbook of mathematical and computational psychology* (pp. 232–254). New York, NY: Oxford University Press.
- Juhasz, B. J., Lai, Y.-H., & Woodcock, M. L. (2015). A database of 629 English compound words: ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior Research Methods*, 47, 1004–1019.
- Kim, S. Y., Yap, M. J., & Goh, W. D. (2019). The role of semantic transparency in visual word recognition of compound words: A megastudy approach. *Behavior Research Methods*, 51, 2722–2732.
- Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review*, 7, 257–266.
- Kintsch, W. (2007). Meaning in context. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 89–105). Mahwah, NJ: Erlbaum.
- Kuperman, V., & Bertram, R. (2013). Moving spaces: Spelling alternation in English noun-noun compounds. *Language and Cognitive Processes*, 28(7), 939–966.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato’s problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211–240.
- Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics*, 20(1), 1–31.
- Li, B., Drozd, A., Guo, Y., Liu, T., Matsuoka, S., & Du, X. (2019). Scaling word2vec on big corpus. *Data Science and Engineering*, 4, 157–175.
- Libben, G. (2006). Why study compounds? An overview of the issues. In G. Libben & G. Jarema (Eds.), *The representation and processing of compound words* (pp. 1–21). Oxford, UK: Oxford University Press.
- Libben, G. (2010). Compound words, semantic transparency, and morphological transcendence. In S. Olson (Ed.), *New impulses in word-formation (Linguistische Berichte Sonderheft 17)* (pp. 212–232). Hamburg: Buske.
- Libben, G. (2014). The nature of compounds: A psychocentric perspective. *Cognitive Neuropsychology*, 31, 8–25.
- Libben, G. (2017). The quantum metaphor and the organization of words in the mind. *Journal of Cultural Cognitive Science*, 1, 49–55.
- Libben, G., Goral, M., & Baayen, R. H. (2018). What does constituent priming mean in the investigation of compound processing? *The Mental Lexicon*, 13, 269–284.
- Lieber, R., & Štekauer, P. (2009). Introduction: status and definition of compounding. In R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of compounding* (pp. 3–18). New York, NY: Oxford University Press.

- Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78.
- Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review*, *122*, 485–515.
- Marelli, M., & Luzzatti, C. (2012). Frequency effects in the processing of Italian nominal compounds: Modulation of headedness and semantic transparency. *Journal of Memory and Language*, *66*, 644–664.
- Marelli, M., Dinu, G., Zamparelli, R., & Baroni, M. (2015). Picking buttercups and eating butter cups: Spelling alternations, semantic relatedness, and their consequences for compound processing. *Applied Psycholinguistics*, *36*, 1421–1439.
- Marelli, M., Gagné, C. L., & Spalding, T. L. (2017). Compounding as abstract operation in semantic space: A data-driven, large-scale model for relational effects in the processing of novel compounds. *Cognition*, *166*, 207–224.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). Efficient estimation of word representations in vector space. [arXiv:1301.3781v3](https://arxiv.org/abs/1301.3781v3).
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems (NIPS) 2013* (pp. 3136–3144). Red Hook, NY: Curran Associates.
- Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, *38*(11), 39–41.
- Mitchell, J., & Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science*, *34*, 1388–1439.
- Murphy, G. L. (1988). Comprehending complex concepts. *Cognitive Science*, *12*, 529–562.
- Murphy, G. L. (1990). Noun phrase interpretation and conceptual combination. *Journal of Memory and Language*, *29*, 259–288.
- Pereira, F., Gershman, S., Ritter, S., & Botvinick, M. (2016). A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. *Cognitive Neuropsychology*, *33*, 175–190.
- Pham, H., & Baayen, H. R. (2013). Semantic relations and compound transparency: A regression study in CARIN theory. *Psihologija*, *46*, 455–478.
- Plag, I. (2006). The variability of compound stress in English: structural, semantic, and analogical factors. *English Language and Linguistics*, *10*, 143–172.
- Rodd, J. M. (2020). Settling into semantic space: An ambiguity-focused account of word-meaning access. *Perspectives on Psychological Science*, *15*, 411–427.
- Rumelhart, D. E., McClelland, J. L., & The PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (volume 1: Foundations). Cambridge, MA: MIT Press.
- Sandra, D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *The Quarterly Journal of Experimental Psychology Section A*, *42*, 529–567.
- Schäfer, M. (2018). *The semantic transparency of English compound nouns*. Berlin, Germany: Language Science Press.
- Schäfer, M., & Bell, M. J. (2020). Constituent polysemy and interpretational diversity in attested English novel compounds. *The Mental Lexicon*, *15*, 42–61.
- Schmidtke, D., Van Dyke, J. A., & Kuperman, V. (2018). Individual variability in the semantic processing of English compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 421–439.
- Smolka, E., & Libben, G. (2017). ‘Can you wash off the hogwash?’—semantic transparency of first and second constituents in the processing of German compounds. *Language, Cognition and Neuroscience*, *32*, 514–531.
- Williams, E. (1981). On the notions “lexically related” and “head of a word”. *Linguistic Inquiry*, *12*, 245–274.
- Wisniewski, E. J. (1997). When concepts combine. *Psychonomic Bulletin & Review*, *4*, 167–183.