

Semantic transparency is not invisibility: A computational model of perceptually-grounded conceptual combination in word processing

Fritz Günther¹, Marco Alessandro Petilli¹, & Marco Marelli^{1,2}

¹University of Milano–Bicocca, Milan, Italy

²NeuroMI, Milan Center for Neuroscience, Milan, Italy

Note: This is the author’s preprint version of the article (date: 28. 01. 2020). The final article is accepted for publication in the *Journal of Memory and Language*.

Previous studies found that an automatic meaning-composition process affects the processing of morphologically complex words, and related this operation to conceptual combination. However, research on embodied cognition demonstrates that concepts are more than just lexical meanings, rather being also grounded in perceptual experience. Therefore, perception-based information should also be involved in mental operations on concepts, such as conceptual combination. Consequently, we should expect to find perceptual effects in the processing of morphologically complex words. In order to investigate this hypothesis, we present the first fully-implemented and data-driven model of perception-based (more specifically, vision-based) conceptual combination, and use the predictions of such a model to investigate processing times for compound words in four large-scale behavioral experiments employing three paradigms (naming, lexical decision, and timed sensibility judgments). We observe facilitatory effects of vision-based compositionality in all three paradigms, over and above a strong language-based (lexical and semantic) baseline, thus demonstrating for the first time perceptually grounded effects at the sub-lexical level. This suggests that perceptually-grounded information is not only utilized according to specific task demands but rather automatically activated when available.

Keywords: Morphological Processing; Conceptual Combination; Embodied Cognition; Deep Learning; Distributional Semantics; Compound Words

A fascinating competence of humans is the ability to combine familiar elements into new, complex ones. For example, by putting a small house on a boat, we can create a houseboat. In fact, we can even perform such a combination on a purely cognitive, conceptual level – which can lead to the concrete implementation of such a combination, but doesn’t need to. Additionally, as can be seen from the “houseboat” example, we are also immediately able to refer to these combinations using linguistic expressions (in this case compound words) and thus to communicate the novel idea to other people.

Therefore, compound words – combinations of two existing words forming a new word, such as *houseboat*, *airport*, or *clickbait* – are often discussed as the linguistic

counterpart to conceptual combination (see Murphy, 2002; Ran & Duimering, 2009; Thagard, 1984), and indeed as the most fundamental combined expressions from which other complex linguistic forms have evolved (Jackendoff, 2002; Libben, 2014). However, such complex linguistic expressions can only be useful if the principle of compositionality – that the meaning of a complex expression can be derived from its constituents and the structure by which they are combined (Frege, 1892) – holds to at least some degree, and thus if the constituents are informative of the intended meaning (Costello & Keane, 2000, based on Grice, 1975). Of course, there is no principled reason stopping us from using completely new expression such as *syllip* to refer to a house on a boat, or even to call it *olivegrass* – after all, linguistic symbols should be largely arbitrary (de Saussure, 1916). However, using the label *houseboat*, and hence adhering to the principle of compositionality, comes with quite obvious advantages in terms of comprehension and communication. Due to this, the use of compositional expressions naturally evolves as a means of communication in scenarios involving complex stimuli (Franke, 2016; Kirby, Cornish, & Smith, 2008). Thus, compounds are not just arbitrary con-

This work was supported by a Research Fellowship (no. 392225719) from the German Research Foundation (DFG), awarded to Fritz Günther, and by grant 2017-1633 from Fondazione Cariplo-Regione Lombardia, awarded to Marco Marelli. Datasets and analysis scripts for this study are available at the Open Science Framework (<https://doi.org/10.17605/OSF.IO/KMRV7>).

catenations of familiar words, but inherently compositional expressions.

From a purely efficiency-based perspective however, it could be argued that a listener or reader doesn't necessarily have to actively compose the meaning of every compound on the basis of its constituents: After repeatedly encountering words such as *houseboat* or *clickbait*, one can eventually form separate whole-word representations for these words and understand them directly (Sandra, 1990; Schreuder & Baayen, 1995). In fact, for semantically opaque compounds such as *ladybird* or *windfall*, these whole-word meanings differ dramatically from their compositionally-obtained meanings (for an overview on semantic transparency/opacity, see Schäfer, 2018). However, from a processing perspective, it still makes sense to immediately initiate a compositional process whenever a compound is encountered (see Chamberlain, Gagné, Spalding, & Lõo, 2019; Günther & Marelli, 2019a). Assuming that the main purpose of language is to convey meaning, language processing would be geared towards *understanding* the linguistic stimuli we encounter (Libben, 2014): Before the whole-word lexical entry has been accessed, one cannot know whether the compound is familiar or not – and thus, whether there even is such an entry (see El-Bialy, Gagné, & Spalding, 2013). The same argument can be made concerning semantic transparency (one cannot know in advance whether the processed complex word will be transparent or opaque; Rastle & Davis, 2008) – and since the distribution of semantic transparency leans heavily towards the transparent side (Gagné, Spalding, & Schmidtke, 2019), the product of a compositional process will be informative of the intended meaning in the vast majority of cases (Rastle & Davis, 2008). As argued by Libben (2006, 2014), compound processing is aimed at maximizing the opportunity to understand the intended meaning in a natural language comprehension context, rather than maximizing efficiency. In this perspective, the processing objective is achieved by immediately initiating a compositional process, rather than delaying it until it is necessarily required.

In line with these theoretical points, several empirical studies have found that the ease of meaning-composition affects compound processing: Libben, Gibson, Yoon, and Sandra (2003) found that compounds whose meaning was rated as predictable from their constituents were processed faster than non-predictable compounds. Marelli and Luzzatti (2012) collected ratings on the semantic transparency of Italian compounds, and found that only ratings on their compositionality (to which degree the meaning of the compound word can be predicted from its constituent meanings), but not on their constituent-compound semantic relatedness, predicted processing times. While both semantic relatedness and compositionality describe the semantic relation between a compound and its constituents, there is a fundamental difference between the two variables: Related-

ness refers to the semantic similarity between the constituent meanings and the whole-word, lexicalized compound; on the other hand, compositionality refers to the active role of constituent meanings in a compound-driven combination process. One consequence of this is that *novel* compounds can be described in terms of compositionality, but not relatedness (Günther, Marelli, & Bölte, in press; Günther & Marelli, 2016, in press). On a psychological level, while relatedness conceptualizes constituent and compound meanings as separate, stored representations in memory, the compositional approach considers the compound meaning as the result of an active process combining the constituents. More detailed discussions on the distinction between the two approaches are provided in Günther and Marelli (2019a); Günther et al. (in press), and Marelli and Luzzatti (2012).

Using a computational model to characterize compositionality, Günther and Marelli (2019a) found that lexical decisions on English compounds are faster in cases where the constituents are more easily integrated into a combined meaning. At the same time, the semantic relatedness between the constituents and the lexicalized, whole-word compound meanings did not affect processing times. In a more recent study, Günther et al. (in press) found that this pattern also holds for lexical decisions in German, and against word-like nonwords (such as *Flughafan* (*airport*) or *Knotenpferd* (*knothorse*)). In fact, the latter class of nonwords are, for all intents and purposes, novel compound candidates. Accordingly, rejections of such items were slower in cases where the constituents are more easily integrated into a combined meaning (Günther et al., in press; Günther & Marelli, in press). Additionally, Amenta, Marelli, and Crepaldi (2015) found, in an eye-tracking study, that a meaning-composition process already sets in very early in the time course of complex word processing.

These results have been interpreted as reflecting an automatic process of conceptual combination in compound processing (in line with, for example, Gagné & Shoben, 1997; Gagné, 2001; Murphy, 1990; Smith & Osherson, 1984; Spalding, Gagné, Mullanly, & Ji, 2010; Wisniewski, 1997; Wisniewski & Love, 1998; see Ran & Duimering, 2009 for an overview), rather than a purely linguistic composition of lexical meanings. However, concepts are more than “just” mental representations of word meanings, or linguistic representations (Kelter & Kaup, 2012). Our conceptual system is not only shaped by the language input we experience, but also by our sensorimotor (i.e., perceptual and motor) experience (Glenberg & Robertson, 2000). This consideration lies at the core of theories of embodied cognition (e.g. Barsalou, 1999; Barsalou, Santos, Simmons, & Wilson, 2008; Fischer, 2012; Glenberg & Robertson, 2000; Glenberg, 2015; Zwaan & Madden, 2005), which have taken a central place in the debate on concept acquisition and representation. In this view, concepts such as DOG are formed through the

interplay between linguistic experience (hearing or reading the word *dog*) as well as sensorimotor experience (seeing a dog, or hearing it bark) (Zwaan & Madden, 2005). Consequently, in language processing, the linguistic stimuli would act as cues to re-activate this sensorimotor experience, or the representation formed from it. Although language typically encodes many aspects of the perceptual world, and language-based representations can come a long way in approximating this perception-based experience (Louwerse, 2011), such redundancies are surely not perfect, and therefore not having direct access to sensorimotor experience will ultimately result in different conceptual representations and processing (Günther, Dudschig, & Kaup, 2018; Kim, Elli, & Bedny, 2019; Striem-Amit, Wang, Bi, & Caramazza, 2018).

It is therefore too narrow a view to understand conceptual combination purely as a language-based meaning-composition process. As concepts are also shaped by sensorimotor experience, these aspects of the concept representations cannot be neglected when considering the combination process. We thus assume that, when a compound is processed, the concepts related to its constituent meanings are accessed. This operation would involve the re-activation of sensorimotor experiential traces linked to the constituents (Zwaan & Madden, 2005), including representations formed from visual experience. These multi-modal representations, in turn, constitute the building blocks of a conceptual combination process, in which they are combined into a new representation (Spalding et al., 2010). A very intuitive illustration (that has, in fact, heavily inspired the present study) of how perceptually-grounded properties can be routinely involved in an automatic conceptual combination process is provided in Figure 1.

As an illustrative example of how this differs from a purely linguistic meaning composition, consider the case of *swordfish*. From a purely language-based perspective, the meaning of this word is not extremely compositional (see Gagné et al., 2019): The semantic properties of both *swordfish* and *fish*, and the language contexts these words occur in, have very little overlap with those of *sword*. From this perspective, the contribution of *sword* to a combined meaning should therefore not be obvious. However, once visual information is considered, this contribution becomes very clear: Swordfish have a long, flat bill that is shaped like a sword, and by “concatenating” the shape of a sword and a fish, one would create something looking very much like a swordfish.¹

In line with such intuitions, Lynott and Connell (2010) argue that traditional theories of conceptual combination cannot account for the role of sensorimotor information, and propose the Embodied Conceptual Combination (ECCo) model to address and overcome these issues. However, this model in its current state has considerable shortcomings: On the one hand, it is a purely verbal theory, which leaves many open degrees of freedom – which are, in practice, usually

filled out by researcher intuition – when it comes to actually testing it. On the other hand, there are only very sparse direct empirical tests of this model (we are only aware of a single study by Connell & Lynott, 2011, which employs a small item set). In a similar vein, Wu and Barsalou (2009) propose that conceptual combination involves spontaneous perceptual simulation of the combined concepts, and provide evidence from property generation tasks: For example, participants produced similar distributions of properties regardless of whether they were instructed to construct mental images for combined concepts or not. However, this study focused on very explicit, off-line tasks rather than on-line processing. In addition, the result of the conceptual combination process for any given combination of elements has to be predicted based on intuition, since the conceptual representations as well as the combination process are under-defined. Furthermore, both previous accounts have not been linked to word recognition and the processing of morphologically complex words, and thus, the role of perceptually-grounded composition effects at the sub-lexical level (that is, concerning units smaller than an individual word) has been left entirely unaddressed.

In the present article, we address these issues by putting forward a model of perceptually-grounded conceptual combination. This model directly includes vision-based information in its very architecture. Being a fully computationally implemented, entirely data-driven model – rooted in representations induced through (deep) neural networks combined with a learning-based compositional system – it also goes beyond verbal theories and researcher intuitions, instead providing completely quantitative characterizations and predictions, which can in turn be subjected to empirical tests. Our model thus naturally allows us to address theoretical hypotheses such as the one discussed so far: If the automatic compositional effect in compound processing observed in previous studies (Günther & Marelli, 2019a; Günther et al., in press) actually reflects a process of conceptual combination, rather than a purely lexical meaning-composition process, then the ease of combining vision-based representations should lead to processing advantages, over and above what is predicted by language alone.

In the following section, we first describe the model architecture and its components – language-based representations, vision-based representations, and the compositional model framework – in detail. In a next step, we describe the measures derived from this model, which quantify the ease of combining the representations, and the contribution of the constituent elements to the newly combined representation. We then proceed to test the model predictions on four large-scale behavioral datasets of compound processing, employ-

¹This example will later be confirmed quantitatively, once we have established the model put forward in this article (see the *Vision-based Measures of Compositionality* section).

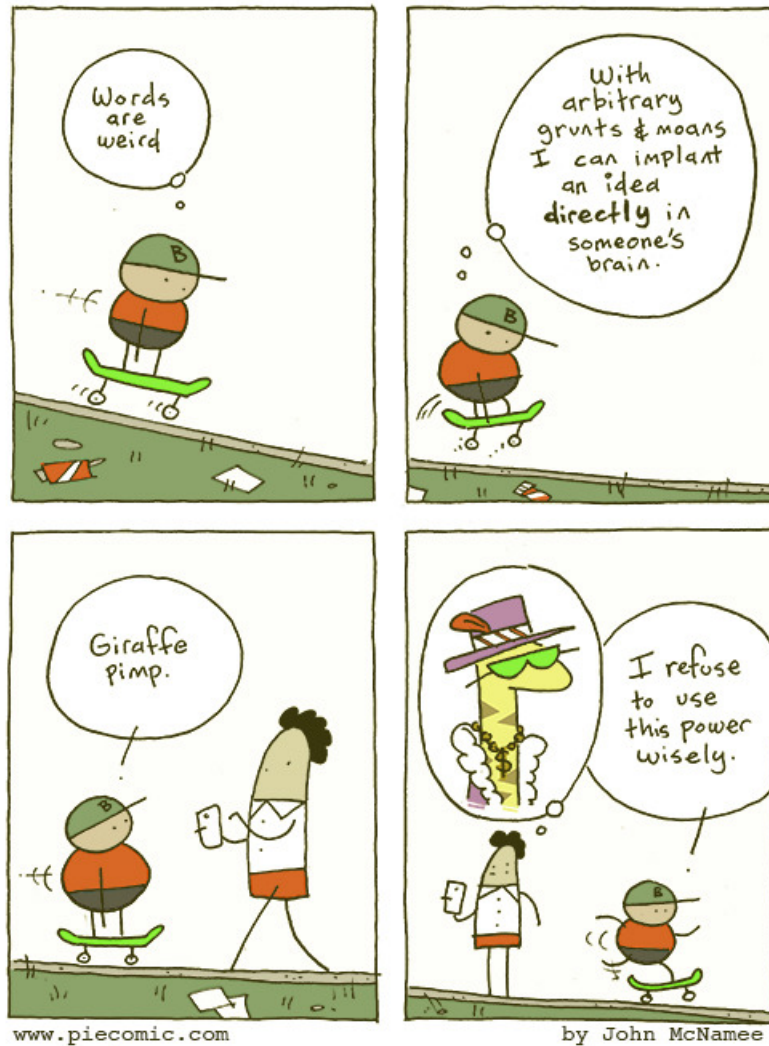


Figure 1. A comic by John McNamee (www.piecomic.com) illustrating on an intuitive level how perceptually-grounded properties can be routinely involved in conceptual combination. All copyrights belong to the artist, who kindly allowed us to use his work here.

ing three different experimental paradigms that require varying degrees of semantic processing (speeded naming, lexical decision, and timed sensibility judgments).

A Computational Framework for Language- and Vision-based Conceptual Combination

Language-based Representations

A graphical illustration of the model architecture and its components is displayed in Figure 2. Language-based representations were obtained via the distributional semantics framework (Landauer & Dumais, 1997; Lenci, 2018; Turney & Pantel, 2010). In distributional semantic models, seman-

tic representations are operationalized as high-dimensional numerical vectors, which are estimated based on the co-occurrence patterns of words in large collections of natural text. The rationale behind this is the *distributional hypothesis*, stating that words with similar meanings occur in similar contexts (Harris, 1954; Firth, 1957), which in its strong version assumes that semantic representations in humans are shaped through such co-occurrence patterns (Lenci, 2008; see also Jenkins, 1954). In numerous studies, it has been shown that distributional semantic models predict human behavior in a large variety of tasks (e.g. Baroni, Dinu, & Kruszewski, 2014; Jones, Kintsch, & Mewhort, 2006; Mandera, Keuleers, & Brysbaert, 2017; Pereira, Gershman, Ritter, & Botvinick, 2016), and there is a wide range of ar-

guments in favor of their plausibility as models of human semantic representation (see Günther, Rinaldi, & Marelli, 2019; Jones, Willits, & Dennis, 2015, for overviews presenting distributional models as theories of semantic memory, and discussing their assumptions and implications).

In the present study, we employed the best-performing word-embeddings model provided by Baroni et al. (2014) to obtain language-based representations. This model was trained on an English ~ 2.8 billion word source corpus (a concatenation of the ukWaC corpus, Baroni, Bernardini, Ferraresi, & Zanchetta, 2009, an English Wikipedia dump, and the British National corpus, BNC Consortium, 2007) using the *cbow* algorithm (with a context window size of 5 words, 400-dimensional vectors, negative sampling with $k = 10$, subsampling with $t = 1e^{-5}$), as implemented in the *word2vec* toolkit (Mikolov, Chen, Corrado, & Dean, 2013; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). The *cbow* algorithm estimates word embeddings (i.e., distributional vectors) as the activation values of the hidden layer of a one-layer neural network model, aimed at predicting a target word from the words in a pre-defined context window (see the upper-left part of Figure 2). Thus, we can in principle derive language-based representation for *all* words occurring in a language via the *cbow* model. In several works, the *cbow* model has been identified as a psychologically plausible model for the acquisition of semantic representations (Hollis, 2017; Mander et al., 2017). In order to obtain reliable word embeddings, we only considered words with a frequency larger than 50 in the source corpus. Examples for language-based neighborhoods (i.e., the most similar word embeddings for a given word) are displayed in Table 1 (left-hand part).

Vision-based Representations

Vision-based representations were obtained from a deep convolutional neural network model, as used in computer vision (Krizhevsky, Sutskever, & Hinton, 2012). Such models are originally trained to predict an image label from a vector representation encoding the pixel-based RGB values of the respective image (see the upper-right part of Figure 2), and have reached impressive levels of performance in this task (Chatfield, Simonyan, Vedaldi, & Zisserman, 2014; Krizhevsky et al., 2012). Furthermore, representations obtained from such models have been validated as measures of visual similarity (Petilli, Günther, Vergallito, Ciaparelli, & Marelli, 2019), and it has been shown that they closely correspond to human intuitions (Bracci, Ritchie, Kalfas, & de Beeck, 2019; Lazaridou, Marelli, & Baroni, 2017; Phillips et al., 2018; Zhang, Isola, Efros, Shechtman, & Wang, 2018).

The starting point for such models is a set of labeled images. In our study, these were obtained from the widely-used ImageNet database (Deng et al., 2009), which adopts the WordNet category structure (Miller, 1995). For

each word in our compound dataset (i.e., the compounds as well as their constituents, see below), a set of 100 to 200 images (depending on the number available) was retrieved by means of ImageNet labels. In cases where a word was used as a label for more than one ImageNet category, the category including more images was selected (see also Petilli et al., 2019). Note that ImageNet “only” contains about 33,000 different image categories – far fewer than the words included in our language-based model – and thus there are many words for which we can derive language- but not vision-based representations. This is not necessarily a technical shortcoming of the specific database, but reflects the actual properties of the considered concepts: For many words where visual experience is lacking (e.g., abstract words), there are principled reasons why a vision-based representation does not exist (see Borghi et al., 2017).²

Vision-based representation vectors for these images were then induced by feeding them to a pre-trained eight-layer deep convolutional neural network (the VGG-F model; Chatfield et al., 2014), as implemented in the MatConvNet Matlab toolbox (Vedaldi & Lenc, 2015). In a recent large-scale evaluation study, Zhang et al. (2018) found that perceptual similarity measures derived from these VGG models outperform a large variety of other models, and come very close to mimicking actual human behavioral data. As established in previous studies (e.g. Lazaridou et al., 2017), we used the activation values of the 4,096-dimensional second-to-last layer of the network (which captures complex, high-level gestalt representations of the images; LeCun, Bengio, & Hinton, 2015; Zeiler & Fergus, 2014) as vision-based representations.

However, at this point, we still have multiple vision-based vectors for each word – one for each of the 100 to 200 images forwarded to the network. From the set of these vectors, we estimated visual prototype vectors as the centroid of all image vectors obtained for a given word (Petilli et al., 2019) which didn’t deviate too far from the median activation value (interquartile ranges over 1.5; Ratcliff, 1993). Examples for vision-based neighborhoods (i.e., images associated with the most similar visual prototype vectors for given words) are displayed in Table 1 (right-hand part).

In order to make the obtained vision-based vectors more computationally manageable, we reduced the original dimensionality of $d = 4,096$ to $d' = 300$ using Singular Value Decomposition (SVD; Martin & Berry, 2007), as implemented in the DISSECT toolkit (Dinu, Pham, & Baroni, 2013). This did not reduce the informativity of the vision-based representations: The original neighborhoods of the prototype vectors as well as the cosine similarities between

²Note that the argument can also run the other way: There can be visual representations for which we have no corresponding linguistic expression, and these cannot be captured by either of the models employed here.

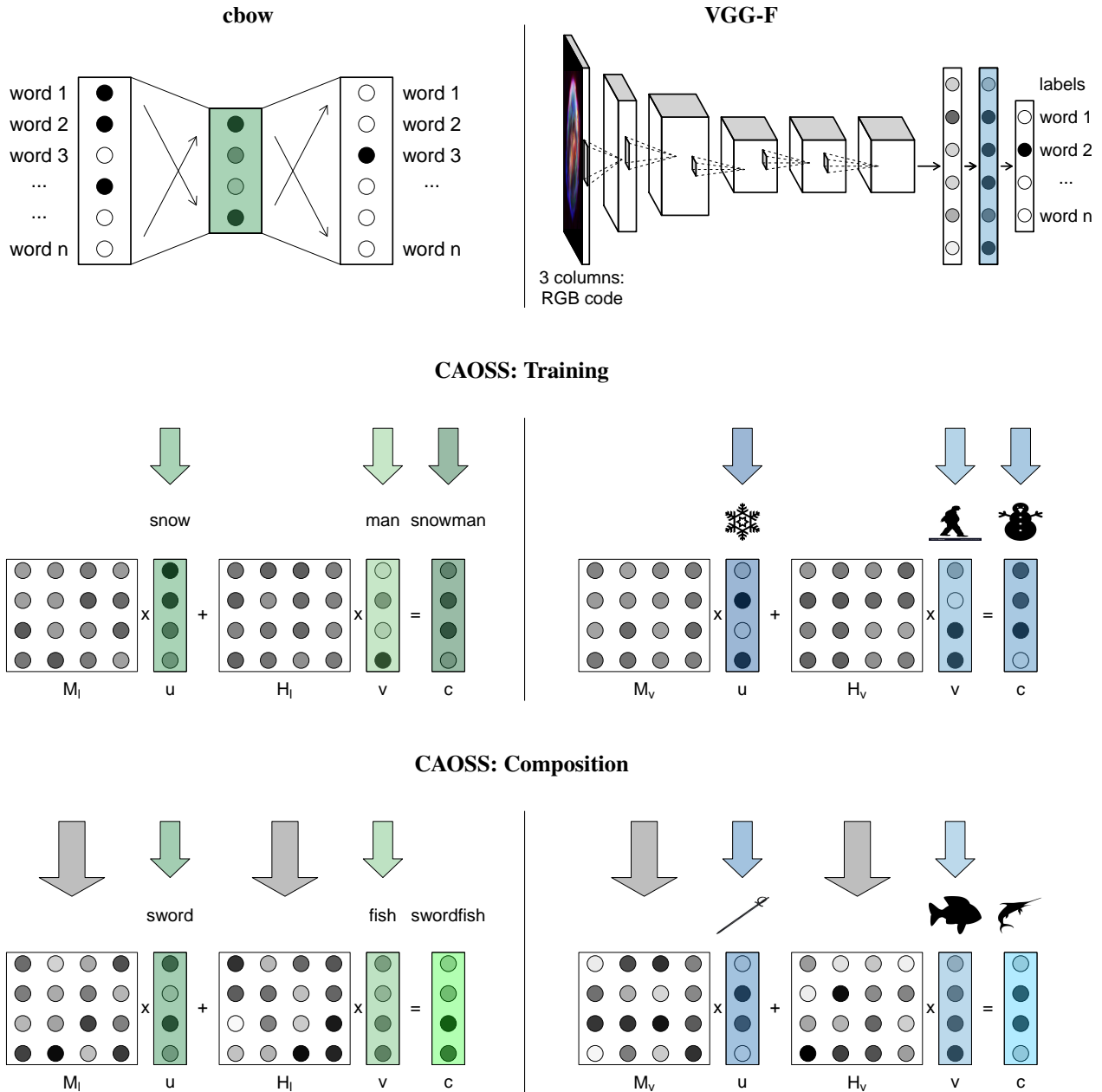


Figure 2. Graphical illustration of the workflow of the models applied here. Vector representations obtained from the word2vec model (cbow algorithm) and the VGG-F model, respectively, are used to train two separate CAOSS models (language- and vision-based). The trained weight matrices are then used to derive vector representations for combinations of constituent vector representations.












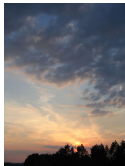



the target and these neighbors were almost perfectly maintained after applying dimensionality reduction. On average, 19.9 out of the 20 closest neighbors to a word were identical, and the average rank correlation between their positions in the neighborhood was $r_s = .99$.

Compositional Model

Using these (language- and vision-based) representations, we trained two compositional models – aimed at predicting compound representations from their constituent representations – using exactly the same compositional architecture (CAOSS; Marelli, Gagné, & Spalding, 2017; see also Guevara, 2010). This model estimates a compositional

Table 1

Examples of language-based (*cbow*) and vision-based (*VGG-F*) neighbors, all included within the 20 most similar representations in the respective model. All reported images are the closest images to the visual prototype representations.

word	neighbors (language-based)	image	neighbors (vision-based)		
<i>orange</i>	yellow, juice, orange-red, khakis, pekoe*		lemon		
			pepper		
					
			grapefruit	tomato	
<i>wolf</i>	werewolf, hellhounds, hyenas, fenris, jackal		fox		
				bobcat	
				wolfhound	
				lion	
<i>sun</i>	moon, sunlight, pleiades, sunshine, moonset		sky		
				fireball	
				rainbow	
				cloud	

compound representation c as

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

, with u and v being the n -dimensional constituent representations, and M and H being a single set of constituent-specific $n \times n$ weight matrices. These matrices are estimated from a training set of existing compounds via least-squares regression, so that the compound representations in the training set

are, on average, best predicted by Equation 1 (see Figure 2). These matrices M and H can then be applied to any new combination of left-hand constituents (the modifier of English compounds) and right-hand constituents (the head of English compounds; Williams, 1981), in order to compute updated, position-specific as-constituent meanings (Marelli et al., 2017; for an in-depth discussion on these as-constituent meanings and their psychological status, see Libben, 2014).

Depending on the specific input vector as well as the specific weights in this matrix, either as-constituent meaning can be very similar to or very different from the original free-word meaning of the constituent: For example, a matrix multiplication with $M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ will not change the input vector $u_1 = (2, 2)$ at all ($M \cdot u_1 = (2, 2)$), but $u_2 = (10, 1)$ will be changed dramatically ($M \cdot u_2 = (1, 10)$). As a result, the outcome of this meaning-updating process will crucially depend on the properties of both the word involved in it and the procedure itself. In the final step, these as-constituent meanings are then added together to obtain a compositional representation for the resulting compound (see Figure 2).

The relatively simple CAOSS model was shown to capture relational effects in the processing of novel compounds (Marelli et al., 2017; see Gagné, 2001; Gagné & Shoben, 1997), compositional effects in the processing of existing and novel compounds (Günther & Marelli, 2019a, in press; Günther et al., in press), and even to predict compound meanings across languages (Günther & Marelli, 2018). Furthermore, it has been shown to outperform a wide range of other possible compositional models (for example, simple vector addition) in a large-scale study (Dima, 2015).

For the language-based representations, the training set consisted of 5,988 compounds: 2,637 hyphenated compounds such as *singer-songwriter*, and 3,351 closed-form compounds such as *airport*, collected from the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995), the English Lexicon Project (Balota et al., 2007), and the compound database by Juhasz, Lai, and Woodcock (2015). For the vision-based representations, the training set consisted of the subset of 388 compound words for which complete compound-constituent sets were available in ImageNet (Deng et al., 2009). Thus, the vision-based model was trained on considerably fewer examples than the language-based model. In addition, since there are many compounds for which both constituents but not the compound itself are available as ImageNet labels, the vision-based composition model is only trained on a subset of the available compounds, while the language-based model is trained on all of them.

Once the models were trained, we applied the estimated CAOSS weight matrices (two sets of M and H matrices, either language-based or vision-based) to induce compositional vectors for all compounds in our datasets for which both constituent representations were available (see below). These computations (training and composition) were performed using the DISSECT toolkit (Dinu et al., 2013). Note that, due to the differences in the training procedure, the vision-based model has to extrapolate from a smaller training set to obtain its compositional representations, and has to do so also for items outside the training set. Examples of predictions of the visual composition model for items that were not included in the training set (i.e., items that are completely

new to the model, and for which it has to rely on the compositional process it learned during training), are displayed in Figure 3.

Method

In the present study, we investigated an item set of 736 existing target compounds. These were selected as all compound words (a) for which a compositional vision-based representation could be obtained (i.e., vision-based representations were available for both constituents), and (b) which were included in each of three large datasets that constitute the empirical basis for the present study (described below). These datasets include data from three different tasks that require different degrees of semantic processing: A naming task, a lexical decision task, and a timed sensibility judgment task. We investigated these different scenarios to ensure the robustness of our model predictions over different task demands, in which semantic processing is required at various degrees. In fact, evaluating potential cross-task dissociations is currently considered one of the main challenges faced by embodied cognition research (Ostarek & Huettig, 2019), where experiments typically rely on paradigms in which sensorimotor processes are explicitly probed. In this perspective, evaluating our model estimates under various experimental conditions will help establishing the routine application of the proposed vision-based compositional process.

Experiment 1: Naming Task

The naming dataset was obtained from the English Lexicon Project (Balota et al., 2007) megastudy. This database contains aggregated naming response times for 40,481 different words, the vast majority of which are not compound words. This data was collected from 444 participants, each of whom was presented with 2,500 words (25 participants per item). In this task, participants saw the word stimuli on a screen and were instructed to read them out aloud as fast and as accurately as possible. The response time was measured as the onset of their response, as recorded with a voice key. The naming task, in principle, could be performed by mapping graphemes to phonemes (a shallow processing task; Barsalou et al., 2008). With a completely transparent orthography, naming could be performed by a foreign speaker who does not even know whether the presented words exist in the target language; thus, this task can in principle be performed without accessing any lexical representation. Certainly, such lexical access may be helpful for naming in a language with opaque orthography such as English, but semantic and conceptual processing remains, in principle, largely unnecessary. Accordingly, previous research identified only limited semantic effects in the naming task (Balota et al., 2007; Hodgson, 1991; Lucas, 2000).

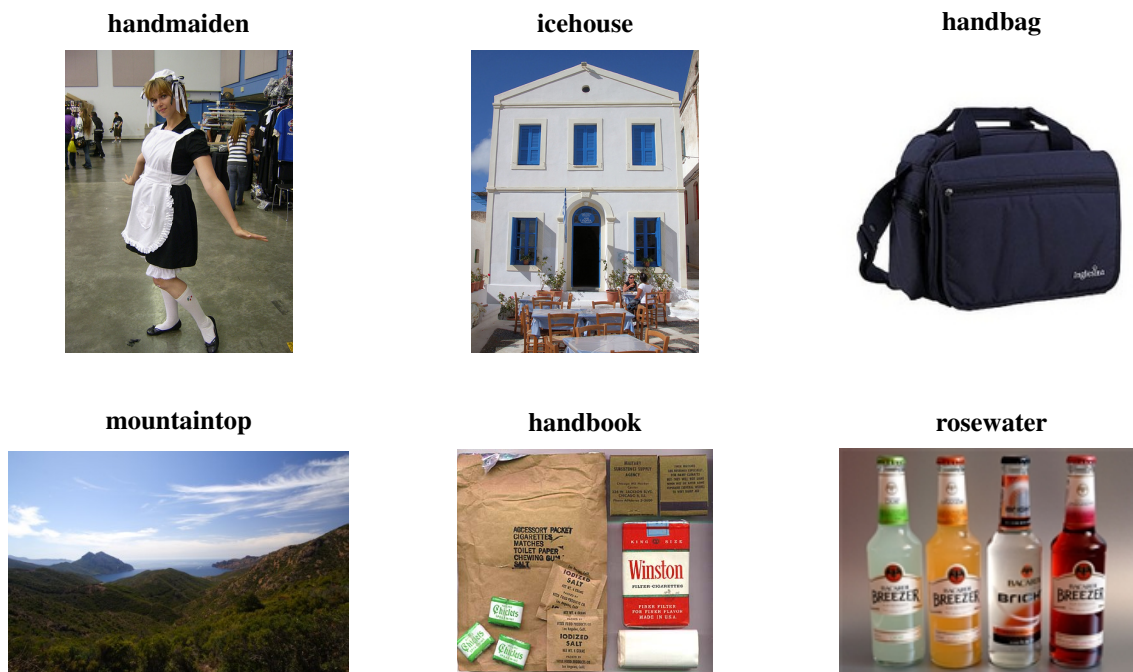


Figure 3. Illustration of (the most similar images to) model predictions of the visual composition model. All these items (displayed in boldface above the images), are not part of ImageNet, and therefore no visual representations can be derived directly using the VGG-F model.

Experiments 2a and 2b: Lexical Decision Tasks

The lexical decision dataset of Experiment 2a was also obtained from the English Lexicon Project (Balota et al., 2007). Aggregated data for 40,481 words was originally collected from 815 participants, each of whom was presented with 1,700 nonwords (created by changing a letter for each of the 40,481 words) and 1,700 target words (34 participants per item). As a supplementary analysis to Experiment 2a, in Experiment 2b we employed a second lexical decision dataset from the British Lexicon Project (BLP; Keuleers, Lacey, Rastle, & Brysbaert, 2012), in order to validate the results of Experiment 2a with data collected from British participants (since with ukWaC and BNC, our source text corpus consists in large parts of British-English documents). The BLP includes lexical decision data for 14,365 words and the same number of nonwords, collected from and aggregated over 78 participants, who each were presented with half of the items (resulting in 38 or 40 participants per item). As a consequence, the BLP only contains data for a subset of 532 out of our 736 target compounds.

In both experiments, participants saw the stimuli on a screen and had to decide – as fast and accurately as possible – for each letter string presented to them whether it was an existing English word. Although the lexical decision task, in principle, only requires lexical access, and not necessarily semantic processing, previous research has es-

tablished that semantic effects do influence lexical decision response times (Amenta, Marelli, & Sulpizio, 2017; Balota et al., 2007; Günther & Marelli, 2019a; Lucas, 2000), although these effects can depend on specific task settings such as the choice of nonwords (Barsalou et al., 2008; James, 1975).

Experiment 3: Timed Sensibility Judgment Task

The timed sensibility dataset was collected by the authors of the present study, and contains data from 145 participants, each of whom was presented with 500 items (22 – 26 participants per item) in a web-based crowdsourcing study (Buhrmester, Kwang, & Gosling, 2011; de Leeuw, 2015; de Leeuw & Motz, 2016). The original dataset contains 1,499 different compound words – all of which are included in the English Lexicon Project – and as many compound “nonwords”. The dataset, along with a detailed methodological description is publicly available (Günther & Marelli, 2019b), via the Open Science Framework (Foster & Deardorff, 2017) at <https://doi.org/10.17605/OSF.IO/7KYNQ>. For this task, the “nonwords” were created by re-combining existing compound constituents into non-existing compounds (i.e., words not observed in the ~ 2.8 billion word language corpus presented earlier, such as *nodemother* or *asylumhiker*). Participants were instructed to decide – as fast and accurately as possible – whether the words presented to them had a sensible interpretation. Response times under 100 ms and

over 5,000 ms were excluded from the data before it was aggregated over participants. Since the timed sensibility judgment task explicitly requires judgments on the meaning of the stimuli, and therefore semantic access to a meaning representation, processing times are expected to be largely influenced by semantic effects (Connell & Lynott, 2013; Estes, 2003; Gagné, 2000).

Linguistic Baseline Measures

The focus of the present study lies on investigating the impact of a vision-based conceptual combination process on response times when processing compound words. Since all our experiments employ inherently linguistic tasks (participants read words and have to respond to them, without any instruction or requirement for mental imagination or visual simulation), we consider as a baseline a variety of language-based lexical and semantic parameters that are known to influence the processing of compound words (Günther & Marelli, 2019a; Kuperman, Bertram, & Baayen, 2008; Kuperman, Schreuder, Bertram, & Baayen, 2009). All language-based measures were derived for the whole set of 736 target compounds.

Compound length was defined as the number of letters in a compound. Word frequency measures – modifier (left-hand constituent) frequency, head (right-hand constituent) frequency, and compound frequency – were obtained from the same ~ 2.8 billion word source corpus from which the language-based representations were derived. All frequencies were logarithmized when entered in any analysis. Modifier and head family sizes (the number of compound types sharing the respective constituent) were obtained from the 5,988-word training set for the language-based CAOSS model, which our training approach assumes to be a representative set of compounds in the source corpus. The three language-based compositional measures were computed as the cosine similarity between the language-based compositional meaning of the compound and (i) the modifier meaning (modifier composition), (ii) the head meaning (head composition), and (iii) the whole-word compound meaning (compound compositionality), see Table 2 and Figure 2 (lower-left part).

Vision-based Measures of Compositionality

We computed two vision-based measures of compositionality as the cosine similarity between the vision-based compositional representation of the compound and (i) the vision-based prototype representation of the modifier (visual modifier composition), and (ii) the vision-based prototype representation of the head (visual head composition), see Table 2 and Figure 2 (lower-right part). These vision-based measures were obtained for the 736 target compounds. Out of these 736 items, we then excluded 10 items for which both visual modifier and head composition had values below -.2, be-

cause strong negative cosine similarities are not interpretable (McNamara, Cai, & Louwerse, 2007).

The correlations between all semantic variables (vision- and language-based) are displayed in Table 3. Notably, for the vision-based model, these two composition measures are highly correlated ($r = .72$) – which is not the case for the language-based model ($r = -.13$; see Table 3). Thus, the notion of “partial transparency” (Libben et al., 2003; Zwitserlood, 1994) seems to be far less relevant for vision-based representations, where compositionality appears to be a rather unidimensional construct. At this point, we can only speculate as to why this is the case. One possibility is that linguistic meanings “react more strongly” to asymmetries: The word usage patterns – and thus, according to the distributional hypothesis (Harris, 1954; Lenci, 2008), the meaning – of *swordfish* will be very similar to *fish* but not to *sword*; similarly, the usage pattern of *stairwell* will be very similar to *stair* but not to *well*. However, the visual representation of *swordfish* still shares many visual features with a sword, and that of a *stairwell* with a *well*, even if this similarity is not reflected in the way we talk about these concepts. Additionally, as is evident in the *well* example, the vision-based model might be less susceptible to homonymy and polysemy (which can dilute language-based similarities) – we only have visual representations for the object-related meanings of words, not for all the other possible meanings.

Thus, since visual modifier and head composition appear to measure very similar constructs, we subsumed these variables in the single parameter *visual composition*, defined as the average of the two measures, to avoid collinearity-related issues such as suppression in our statistical analyses. Correlations between language- and vision-based measures of compositionality were quite low ($r = .17$ for both the correlation between language- and vision-based modifier composition and head composition), indicating that the language- and the vision-based (composition) models indeed capture different pieces of information. Note that no visual equivalent to compound compositionality was computed, since no images (and therefore no observed, “whole-word” vision-based representations) are available in ImageNet for 352 out of 736 compounds in the dataset.

Having established these measures, we can reconsider the *swordfish* example: Its language-based modifier composition (.40) has a percentile value of 44.8 within the set of 726 compounds for which vision-based representations are available, while its language-based head composition (.57) has a percentile value of 94.6. On the other hand, its vision-based modifier composition (.89) has a percentile value of 90.9, while its vision-based head composition (.82) has a percentile value of 63.4. As can be seen from this example, integrating *sword* into the combined concept of *swordfish* is indeed easier when considering visual information rather than linguistic data.

Table 2

Language- and vision-based measures of semantic compositionality. All similarities are defined as cosine similarities between the respective vector representations (language-based word embeddings or vision-based prototype vectors). Abbreviation: *cmpd.* for compound, *vis.* for visual.

type	measure	definition	examples
language-based	modifier composition	cos(modifier, compositional <i>cmpd.</i>)	low: <i>reindeer, wormwood</i> high: <i>millstone, roadhouse</i>
	head composition	cos(head, compositional <i>cmpd.</i>)	low: <i>stairwell, hornbill</i> high: <i>firewater, silverfish</i>
	compound compositionality	cos(whole-word <i>cmpd.</i> , comp. <i>cmpd.</i>)	low: <i>windfall, clubfoot</i> high: <i>saucepan, sugarcane</i>
vision-based	vis. modifier composition	cos(vis. modifier, vis. compositional <i>cmpd.</i>)	low: <i>deerskin, fingerprint</i> high: <i>tramcar, hillside</i>
	vis. head composition	cos(vis. head, vis. compositional <i>cmpd.</i>)	low: <i>witchdoctor, beetroot</i> high: <i>shipwreck, soybean</i>

Table 3

Correlations between measures of language- and vision-based compositionality.

	head comp.	comp. compos.	vis. mod. comp	vis. head comp	vis. composition
mod. comp.	-.13	.18	.17	-.01	.08
head comp.		.16	.03	.17	.11
comp. compos.			.13	.14	.14
vis. mod. comp				.72	.92
vis. head comp					.94

Results

Having computed all required measures, we test for each of the three tasks investigated here (naming, lexical decision, timed sensibility judgments) the hypothesis that differences in vision-based compositionality (as measured by the *visual composition* metric) predict the performance in behavioral tasks over and above language-based effects – including lexical (length, logarithmic constituent and compound frequencies, family sizes) as well as semantic variables (Table 2). This analysis therefore concerns effects *within* the set of compounds for which visual composition values can be derived. An additional analysis demonstrating that there are also systematic processing differences *between* these compounds and non-visually-represented ones – in line with a general concreteness/imageability effect in compounds (Feldman, Basnight-Brown, & Pastizzo, 2006; Schmidtke & Kuperman, 2019; see Paivio, 1966, 1986) – is provided in Supplementary Material A.

The testing procedure is displayed in Figure 4. For each task, we estimated a baseline Linear Mixed Effect Model (Baayen, Davidson, & Bates, 2008) to predict the dependent variable, using the R packages *lme4* (Bates, Mächler, Bolker, & Walker, 2015) and *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017). As predictors, this model contained all language-based lexical and semantic measures, as well as random intercepts for the compound modifiers and

heads. These random effects were included to capture the partial repeated-measure structure of the data resulting from repeated modifiers and heads, and to account for item variability without introducing an idiosyncratic effect for each item (since we have exactly one observation per item in each analysis). We then estimated another model, containing the very same parameters plus an additional parameter for visual composition. The resulting model was then compared to the baseline model using a likelihood-ratio test, to test whether the inclusion of the additional parameter significantly improved the model fit.

Obviously, the resulting models will contain many non-predictive parameters from the baseline models, some of which are correlated to other variables, rendering the parameters of any resulting model difficult to interpret. In order to obtain interpretable final models, we removed all non-significant effects of the final models in a step-wise backward selection procedure (Kuznetsova et al., 2017).

The raw data and R scripts for all analyses reported here are available at the Open Science framework (<https://doi.org/10.17605/OSF.IO/KMRV7>).

Response Times

Response times were log-transformed for the analyses (Baayen & Milin, 2010). In all four experiments, including a visual composition parameter significantly improves

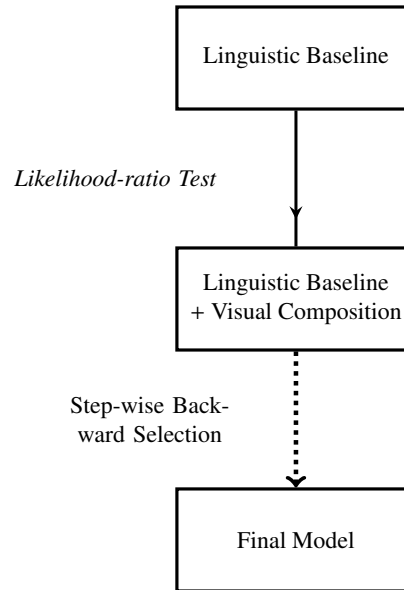


Figure 4. The testing procedure applied for the analyses presented here.

the baseline model (naming: $\chi^2(1) = 4.49$, $p = .034$; lexical decision: $\chi^2(1) = 4.35$, $p = .037$ in the ELP, $\chi^2(1) = 4.37$, $p = .037$ in the BLP; timed sensibility judgments: $\chi^2(1) = 6.13$, $p = .013$). The final model parameters predicting response times (after elimination via backwards selection) for all three tasks are displayed in Table 4³. Variance inflation due to collinearity was not an issue for any reported model (variance inflation factors for all variables between 1.01 and 1.17, estimated using the R package *usdm*; Naimi, Hamm, Groen, Skidmore, & Toxopeus, 2014). The visual composition parameter in all resulting models is negative, indicating a facilitatory effect of vision-based compositionality: the more easily the denoted object can be visually composed, the faster are the observed response times to the word.

In a follow-up analysis to compare task effects, we then estimated an overall model for the three tasks (naming, ELP lexical decision, timed sensibility), including as fixed effects all possible interactions between the predictors and a dummy variable encoding the task, as well as random intercepts for modifiers and heads. The BLP data was not included in this analysis, since it contains fewer items than the other tasks. To specifically test if the visual composition effect is different between tasks, we compared this full model with a model that contains the same fixed effect structure except for the interaction between task and visual composition. Removing this interaction did not significantly affect model predictions ($\chi^2(2) = 1.21$, $p = .545$); we therefore have no indication for a task-dependency of the visual composition effect. The parameters of the final model after exclusion of non-significant parameters via backwards selection is provided in Table 5.

Note that, while the study by Günther and Marelli (2019a) observed effects of modifier- and head-composition in a lexical decision task, the same pattern did not emerge in the present study (see Table 4). A post-hoc analysis establishing the conformity between our present results and those of the Günther and Marelli (2019a) study is provided in Supplementary Material A.

Accuracy

We further investigated accuracy data across all four experiments. However, since the tasks are relatively easy, accuracy data is extremely skewed in all of them, with very high or even perfect accuracy for a large majority of items (median accuracies are .92 for the timed sensibility task, .94 for the ELP lexical decisions, .90 for the BLP lexical decisions, and 1.00 for naming). Accuracy values were logit-transformed for this analysis (replacing values of 0.0 with 0.1 and 1.00 with .99, since the logit for these values is infinity.)

The visual composition parameter significantly improves the baseline model in the two lexical decision tasks (ELP: $\chi^2(1) = 5.76$, $p = .016$; BLP: $\chi^2(1) = 5.33$, $p = .021$) and in the timed sensibility judgment task ($\chi^2(1) = 7.94$, $p = .005$). Visual composition did not predict the accuracy in the naming task ($\chi^2(1) = 0.10$, $p = .753$), but there is extremely little variance to explain in naming accuracy in the first place (mean = .98, $Q_1 = .96$, median = 1.00).

The final model parameters predicting accuracies (after elimination via backwards selection) for all three tasks are dis-

³Note again that the BLP analysis is based on 532 out of the 726 items in the other analyses.

Table 4

Model parameters – regression weights (t-values) – predicting logarithmic response times in the final models (after elimination of non-significant parameters via backwards selection), for all four analyses. Abbreviations: modifier (mod.), compound (comp.), frequency (freq.). Only the significant effects which remain in the resulting models are displayed ($p < .05$).

type	parameter	naming		LDT (ELP)		LDT (BLP)		timed sensibility	
	intercept	6.74	(124.69)	6.89	(111.52)	6.82	(317.72)	7.11	(124.72)
lexical	length	0.03	(10.10)	0.03	(7.20)			0.02	(4.02)
	comp. freq.	-0.02	(-8.96)	-0.05	(-13.72)	-0.04	(-15.35)	-0.04	(-11.67)
	mod. freq.	-0.01	(-4.17)	-0.01	(-3.69)				
	head freq.	-0.01	(-3.80)						
	mod. family size					-0.001	(-2.11)		
	head family size							0.01	(3.42)
language-based	mod. composition							-0.12	(-2.07)
	head composition							-0.14	(-2.78)
	comp. compos.			0.10	(2.11)				
vision-based	visual composition	-0.04	(-2.11)	-0.05	(-2.08)	-0.05	(-2.53)	-0.06	(-2.37)

Table 5

Model parameters – regression weights (t-values) – predicting log-transformed response times in the overall task-combined analysis (after elimination of non-significant parameters via backwards selection). The ELP lexical decision task serves as a reference level for this analysis. Abbreviations: modifier (mod.), compound (comp.), frequency (freq.), timed sensibility (TS). Only the significant effects which remain in the resulting models are displayed ($p < .05$).

type	parameter	<i>b</i>	<i>t</i>
	intercept	7.04	(102.51)
task	naming	-0.21	(-3.49)
	TS	0.02	(0.40)
lexical	length	-0.03	(-7.45)
	comp. freq.	-0.04	(-12.85)
	mod. freq.	-0.01	(-2.84)
	head freq.	-0.01	(-1.96)
	mod. family size	-0.00	(-0.84)
task : lexical	naming : length	-0.00	(0.16)
	TS : length	-0.01	(-2.20)
	naming : comp. freq.	-0.02	(5.16)
	TS : comp. freq.	-0.00	(-0.90)
	naming : head freq.	-0.00	(-0.79)
	TS : head freq.	0.02	(4.67)
	naming : mod. family size	-0.00	(-1.19)
	TS : mod. family size	0.00	(1.84)
language-based	mod. composition	-0.01	(-2.42)
vision-based	visual composition	-0.01	(-3.00)

played in Table 6. Variance inflation factors for all variables lay between 1.0 and 1.11, indicating no collinearity issues. The visual composition parameter in all resulting models is positive, so the more easily the denoted object can be visually composed, the more likely participants are to give the correct answer in lexical decision and timed sensibility judgments.

In the overall analysis of accuracy data (which was set up analogously to the overall analysis for response times),

we observe an interaction between visual composition and task (removing this parameter leads to significantly worse model fit; $\chi^2(2) = 8.44, p = .015$). This is due to the fact that the visual composition effect is absent in naming accuracies (see Table 7, which displays the parameters of the final model after exclusion of non-significant parameters via backwards selection). However, as already noted, one should be careful in the interpretation of these naming accuracies.

Table 6

Model parameters – regression weights (t-values) – predicting logit-transformed accuracies in the final models (after elimination of non-significant parameters via backwards selection), for all three tasks. Note for parameter interpretation that $\text{logit}(.50) = 0$, and $\text{logit}(.99) = 4.60$. Abbreviations: modifier (mod.), compound (comp.), frequency (freq.). Only the significant effects which remain in the resulting models are displayed ($p < .05$).

type	parameter	naming	LDT (ELP)		LDT (BLP)		timed sensibility	
	intercept	1.59 (14.53)	-1.00 (-3.94)		-1.48 (-7.47)		-0.16 (-0.49)	
lexical	length	-0.02 (-2.26)					-0.06 (-3.22)	
	comp. freq.	0.04 (6.18)	0.13 (9.06)		0.32 (14.17)		0.23 (13.51)	
	mod. freq.	0.03 (3.53)	0.10 (5.36)				0.05 (2.41)	
	head freq.						-0.05 (-2.77)	
language-based	mod. composition		0.66 (2.67)				0.57 (2.18)	
vision-based	visual composition		0.25 (2.37)		0.46 (2.74)		0.33 (2.77)	

Table 7

Model parameters – regression weights (t-values) – predicting logit-transformed accuracies in the overall task-combined analysis (after elimination of non-significant parameters via backwards selection). The ELP lexical decision task serves as a reference level for this analysis. Abbreviations: modifier (mod.), compound (comp.), frequency (freq.), timed sensibility (TS). Only the significant effects which remain in the resulting models are displayed ($p < .05$).

type	parameter	<i>b</i>	<i>t</i>
	intercept	-0.84	(-2.76)
task	naming	1.99	(7.73)
	TS	0.58	(2.25)
lexical	length	0.01	(0.66)
	comp. freq.	0.12	(9.67)
	mod. freq.	0.01	(3.17)
	head freq.	0.03	(1.87)
	mod. family size	0.01	(2.45)
task : lexical	naming : length	-0.04	(-2.37)
	TS : length	-0.08	(-4.76)
	naming : comp. freq.	-0.10	(-6.38)
	TS : comp. freq.	0.10	(6.09)
	naming : head freq.	-0.02	(-1.35)
	TS : head freq.	-0.06	(-4.15)
	naming : mod. family size	-0.004	(-3.55)
	TS : mod. family size	-0.004	(-3.23)
language-based	mod. composition	0.51	(2.85)
vision-based	visual composition	0.22	(2.31)
	naming : visual composition	-0.22	(-2.08)
	TS : visual composition	0.11	(1.03)

Discussion

In the present study, we investigated effects of vision-based compositionality – that is, the ease of combining the visual representations of the objects denoted by compound constituents into a single newly-composed representation – on the processing of compound words. To this end, we considered four large-scale behavioral experiments of compound word processing employing three different tasks: a naming experiment, two lexical decision experiments, and a timed

sensibility judgment experiment. In principle, these tasks require different degrees of semantic processing to be successfully performed. Across tasks, we found that vision-based compositionality predicts compound processing times and accuracies over and above a baseline of language-based measures of semantic compositionality and other lexical variables. Leaving aside the overall extremely high naming accuracies, we find no indication for a difference in the strength of the visual composition effect between different

tasks. This indicates that compound processing entails an automatic compositional process combining the constituent meanings, and further indicates that such composition is also grounded in perceptual experience.

Perceptually-Grounded Conceptual Combination

Critically, the present results characterize this meaning-composition process not as a purely language-based composition of lexical meanings, but rather as a full-fledged conceptual combination process that also relies on the perceptually grounded information available for the constituent concepts (Lynott & Connell, 2010; Wu & Barsalou, 2009). Indeed, the vision-based component even appears to be more adequate to characterize the conceptual combination process, as we find vision-based effects of compositionality even when language-based compositionality effects are relatively weak.

To our knowledge, the present study is the first to systematically examine perceptually grounded effects at a sub-lexical composition level. While semantic effects in the processing of morphologically complex words have received considerable research attention, this research line was so far based on language-centric conceptualizations of semantic representation (e.g., Günther & Marelli, 2019a; Libben et al., 2003; Schmidtke, Van Dyke, & Kuperman, 2018; Marelli & Baroni, 2015; Marslen-Wilson, Tyler, Waksler, & Older, 1994; Rastle, Davis, Marslen-Wilson, & Tyler, 2000; Sandra, 1990; Smolka, Preller, & Eulitz, 2014; Zwitserlood, 1994). However, as indicated by the present results, a wider conceptual-level approach that also considers perception-based representations and processes (Barsalou, 1999; Kelter & Kaup, 2012; Zwaan & Madden, 2005) might be more adequate to investigate semantic effects than an approach narrowed down to lexical-meaning representations.

On the other hand, embodied theories of conceptual combination (Lynott & Connell, 2010; Wu & Barsalou, 2009) are typically concerned with novel multi-word combinations rather than familiar complex words, and focus on explicit interpretations rather than on-line processing. Thus, up to now, the literature on embodied cognition and on morphological processing seem to have completely ignored each other. The most similar study to the one presented here was conducted by Connell and Lynott (2011), who investigated processing times in an interpretation task for 27 novel two-word phrases (similar to the timed sensibility task employed in the present task). However, our study significantly extends upon this study, by employing two other tasks which do not necessarily require semantic processing and are heavily employed in the morphological processing literature (naming and lexical decision), by employing independently-obtained, data-driven measures rather than researcher intuitions, and by testing item sets that are many times larger than the one by Connell and Lynott (2011).

Furthermore, the present study demonstrates perceptually-grounded combination in existing rather than the novel combinations which so far have been the focus of the embodied conceptual combination literature. We thus bring together the research lines on morphological processing on the one hand, and embodied cognition on the other hand, which have so far been completely disconnected. However, their topics of interest have much in common: Both examine how we are able to extract meaning from the arbitrary flow of symbols that is language, and how the form of these symbols is systematically connected to the concepts and mental representations for which it serves as a cue. Further, both fields have already been linked with conceptual combination, which plays a vital role in advancing our conceptual system (El-Bialy et al., 2013; Gagné & Spalding, 2004; Ji, Gagné, & Spalding, 2011; and Lynott & Connell, 2010; Wu & Barsalou, 2009) – although one can argue that this connection is also under-explored from both sides. The present study demonstrates that these research lines naturally complement one another, and that their potential synergy can result in new insights furthering our understanding in all involved fields.

Importantly, none of the tasks employed here required or even encouraged participants to engage in mental imagination or visual simulation. Rather, all three experiments employed completely language-centered tasks, two of which (naming and lexical decisions) do not even, in principle, require any kind of semantic processing. As highlighted by Ostarek and Huettig (2019), these are no trivial conditions for observing effects of sensorimotor activation: These effects are often only found when sensorimotor simulation processes are explicitly encouraged, and there are serious doubts that such processes are automatic (Lebois, Wilson-Mendenhall, & Barsalou, 2015). Nevertheless, our findings are in line with results from previous studies: Petilli et al. (2019) found that the visual similarity between prime-target word pairs (also measured via the VGG-F model, Chatfield et al., 2014) predicted semantic priming effects in a lexical decision task. Günther and Marelli (2019a) observed an automatic effect of semantic compositionality in a lexical decision task, which was interpreted as reflecting conceptual combination; and at the same time, Wu and Barsalou (2009) demonstrated that participants engage in perceptual simulation during conceptual combination, irrespective of whether they were instructed to do so. Considering these findings alongside our present results, we conclude that conceptual combination always involves the combination of perceptually-grounded information. Of course, we assume that this only happens in cases where perceptual experience with the constituents is available.

This is in line with the LASS (language and situated simulation) theory of embodied language processing by Barsalou et al. (2008), which postulates that both lin-

guistic processing and perceptually-grounded simulations always take place during conceptual processing. According to the LASS theory, the relative role of these two components varies greatly depending on the task requirements: In explicitly perceptually-related task, such as the generation of perceptual features for objects, perceptual simulation should play the leading role. However, in tasks that require relatively shallow processing – such as lexical decision or naming – linguistic processing should be sufficient for successful performance and therefore play the leading role, with perceptual simulation hardly being involved. The present results refine this picture, while still adhering to the general reasoning by Barsalou et al. (2008): If a processing step runs automatically (as in the case of conceptual combination processes), it will influence performance in the task irrespective of task demands. This is a common finding across the psychological literature: For example, already the classic study by Stroop (1935) demonstrates that completely task-irrelevant processes (in this case reading and, in fact, semantic access to the respective word meanings) affect task performance, if they are executed automatically.

Implications of the Model Architecture

Being implemented on powerful, data-driven models of linguistic and visual representation (Chatfield et al., 2014; Mikolov, Chen, et al., 2013), our model allows us to test precise, quantitative predictions for effects of vision- and language-based semantic compositionality in the processing of complex words. This enables us to investigate large datasets of items, for which predictions based on human intuition would be extremely laborious to collect, and most likely highly ambiguous or inconsistent for many items. To our knowledge, the present study is the first to put forward a fully implemented, cognitively plausible, and structure-sensitive model of vision-based conceptual combination that is tested on large-scale behavioral datasets. The closest candidate for such a model so far was presented by Pezzelle, Shekhar, and Bernardi (2016), who proposed a compositional model based on simple vector addition. However, these authors only tested the model predictions internally, as the correspondence between the model-predicted compositional vectors with observed vectors for images depicting compound word referents. In addition, on an empirical level, studies in the language domain have demonstrated that a CAOSS-style model outperforms simple vector addition, also when it comes to maximizing the correspondence between model-predicted and observed vectors (Dima, 2015). And on a theoretical level, simple vector addition implies implausible assumptions about compounding (Marelli et al., 2017): For example, differential constituent roles cannot be considered, despite the fact that compounds are inherently asymmetric constructions (Di Sciullo, 2005; a *houseboat* is not a *boathouse*, but $\overrightarrow{house} + \overrightarrow{boat} = \overrightarrow{boat} + \overrightarrow{house}$).

It is also important to note that the vision-based compositional model, although trained on representations induced from images, also builds on linguistic categorizations, since it moves from images that are annotated with linguistic labels – in this case, compounds and their constituents. For example, the training item *houseboat* will be a vision-based representation constructed from images labeled as *houseboat*, which is to be predicted from a representation based on images labeled as *house* and representation based on images labeled as *boat*. Since the training item is a compound word, the model will be fed a structure that it can adapt to: One constituent is in the left-hand position and will be updated through the weight matrix M , and the other is in the right-hand position and will be updated through another weight matrix H . Critically, these positions are not arbitrary in English compounds: In most cases, the word in the right-hand position specifies the object category (a *houseboat* is a kind of *boat*). While it could be argued that this imposes a language-based structure on the vision-based system, we would attribute this structure to a conceptual rather than a linguistic level: A combination of visual objects (or the corresponding vision-based representations) is not just a symmetric blending of visual features, and the visual features are combined differently in a *boathouse* than in a *houseboat*. Training the CAOSS matrices of the visual composition model on compounds allows them to capture this structure, if needed. For example, one could speculate that the right-hand constituent, typically defining the compound category, provides more shape-related information about the resulting combined representation than the left-hand constituent: A *bluebird*, *redbird* and *blackbird* have very similar shapes, and differ mostly in color.

Importantly, our architecture relies on the exact same compositional system (although applied on different input vectors and training sets) to derive language- and vision-based representations for combined concepts (the CAOSS model, Marelli et al., 2017). Thus, while this compositional model is quite simple (it learns a weighted-addition combination function from experience with combinations of the same type, and then applies what it learned to combine new elements), it is flexible enough to handle different types of input derived from qualitatively very different sources. In fact, Günther and Marelli (2019a) already proposed that the CAOSS model is not restricted to distributional word embeddings (on which it was originally implemented), but can be applied for any type of vector-based dimensional representation. Thus, unlike previous theories of embodied conceptual combination (Lynott & Connell, 2010; Wu & Barsalou, 2009) we don't have to assume any qualitative differences in the conceptual combination procedure for different types of input; the architecture proposed here at most warrants gradual differences in the weighting of dimensional attributes.

In a more general context, our approach is based

on the fundamental assumption that the cognitive system utilizes and adjusts to statistical regularities in the environment to construct mental representations and concepts (Günther et al., 2019; Ramscar, Hendrix, Love, & Baayen, 2013). This assumption constitutes the basis for both the language-based model (Landauer & Dumais, 1997; Westbury, 2016) and the vision-based model (Chatfield et al., 2014; Krizhevsky et al., 2012), as well as the compositional system itself (Marelli et al., 2017). This perspective directly parallels core theoretical proposals both in the fields of embodied cognition (Zwaan & Madden, 2005) and morphological processing (Baayen, Milin, Filipović Đurđević, Hendrix, & Marelli, 2011; Günther, Smolka, & Marelli, 2019; Milin, Feldman, Ramscar, Hendrix, & Baayen, 2017; Rastle, 2018), which attribute a central role in shaping the cognitive system and its representations to learning systematic patterns linking linguistic forms on the one hand to meaning representations and concepts on the other hand.

While in the present study we focused on the visual domain to approximate sensorimotor experience, we are of course not implying that visual experience is the only relevant type of such experience. Still, the visual domain is arguably the most prominent one, in our cognitive representations as well as in psychological and cognitive science research. As a result, vast amounts of work have been invested in the field of computer vision over the last years (as of this writing, the study Krizhevsky et al., 2012, alone is cited over 55,000 times on Google Scholar), and models of visual representation are in a state where they produce high-quality results (Zhang et al., 2018). Nevertheless, as discussed in the previous paragraph, our model architecture is flexible enough to perform combinations of any type of vector representation. At the same time, recent studies have started to propose models to obtain auditory (Kiela & Clark, 2015; Lopopolo & Miltenburg, 2015) or even olfactory representations (Kiela, Bulat, & Clark, 2015). As these models start to produce high-quality representations, they can be easily integrated in the architecture proposed here, and other modalities can be considered alongside vision.

Note that, in the model architecture presented here, the language- and vision-based systems essentially run “in parallel”, with two separate compositional systems associated with them. With this, we don’t want to claim that we have separate mental representations for a single concept, and separate combination systems, each informed by a specific modality. Instead, we opted for this approach in order to disentangle the information provided from different input systems, and to test for the specific influence of vision-based information over and above what is available purely from language input. To this end, we computed and employed a variable exclusively encoding this information (visual composition), which allows us to estimate the additional predictive power brought specifically by information from the vi-

sual domain.

In principle, there is also the possibility to combine language- and vision-based vectors to derive a unitary representation for each concept, for example by concatenating them (Andrews, Vigliocco, & Vinson, 2009; Bruni, Tran, & Baroni, 2014) or mapping the two systems onto one another (Lazaridou, Pham, & Baroni, 2015; Lazaridou et al., 2017). However, such approaches come with their own problems: The relative number of dimensions from the language- and the vision-based part is arbitrary, but in a concatenation-based system this parameter will impact our composition model, whose weight matrices capture the influence that every input element has on every output element. Also, while at first appearing cognitively plausible, concatenation-based models cannot effectively account for concepts for which no information is available from one source.

Additionally, in all these models, language- and vision-based representations are acquired from completely independent contexts and training instances. Therefore, to reach a truly realistic model, we advocate for an approach that learns language- and vision-based representations from the very same contexts, in which linguistic and visual experience co-occur together (see Günther et al., 2019, for a theoretical proposal of such a model based on the experiential trace model by Zwaan & Madden, 2005).

Automatic Compositional Processes

The assumption that the effects of compositionality reported here are due to an active meaning-composition process and not to activation of information linked with a stored whole-word representation is in line with earlier empirical results and theoretical arguments: First, compositionality is a property that by definition involves not only the compound representation, but also the constituent representations. In this context, Günther and Marelli (2019a) have demonstrated that the interplay between compound and constituent representations in compound processing has to be understood in compositional terms, since processing times are only explained by the contribution of constituents to the compositional compound representation, and not by their semantic relatedness to the whole-word compound representation.

This still leaves open the possibility that, once constructed, the compositional representation is stored alongside with, or in addition to, the whole-word representation. However, this can severely hinder access to the whole-word representation required for actual comprehension in the case of non-compositional, opaque compounds, and at best not be helpful for highly compositional compounds. Thus, compositional representations are expected to be helpful only in on-line processing, where we don’t know whether we are familiar with a given complex word (El-Bialy et al., 2013) – and in order to understand novel words, engaging in a compositional process is necessary. However, due to this very fact, the meaning-

composition process is initiated and executed in any case: Delaying it would cause a breakdown in understanding in cases where it is actually needed (Libben, 2014), and in the majority of cases where the compound is transparent (we also cannot know in advance whether the word might be opaque; Rastle & Davis, 2008), the compositional process can facilitate access to the whole-word meaning and thus word comprehension. In fact, this assumption is supported by recent studies demonstrating that the compositional semantic effects in novel compound processing mirror those for existing words (Günther & Marelli, in press; Günther et al., in press).

Such an automatic process of conceptual combination necessarily requires that the individual constituent representations have been identified from the compound, which is encountered as a single character string. In the morphological processing literature, several models have been proposed and discussed how this can be achieved (for an overview, see Amenta & Crepaldi, 2012): The activation of morphology-related information can be the result of a meaning-blind, automatic decomposition process that segments any seemingly morphologically-complex string into its potential constituents (Marslen-Wilson et al., 1994; Rastle, Davis, & New, 2004; Rastle & Davis, 2008); however, it can even be considered a by-product of a learned systematic mapping between orthographic cues and lexico-semantic representations (Baayen et al., 2011; Milin et al., 2017). In this context, it has already been argued that the function and purpose of this process is to identify the constituents that subsequently serve as the building blocks for a concept-composition process (this argument is explicit in the study by Rastle & Davis, 2008). This in turn enables speakers to rapidly understand the meaning of novel combinations in natural language (Libben, 2014), which in turn allows us to use such combinations to communicate of new ideas, and to create new words to express a new meaning when we need it (Downing, 1977).

However, with this emphasis on compositional processing we don't want to claim that the whole-word compound representation is irrelevant for the representation and processing of existing compounds. At some point during processing, the whole-word representation clearly is accessed, which is for example demonstrated by the consistently strong compound frequency effect over and above the constituent frequency effects (see Table 4 and Table 6). In addition, the effect of compound compositionality (the similarity between the language-based compositional and whole-word meaning) in the ELP lexical decision data indicates that also the whole-word meaning of the compound is accessed. Furthermore, Günther et al. (in press) investigated semantic effects across multiple tasks, and observed that the whole-word meaning plays a central role for explicit judgments on the compound meaning (see also Marelli & Baroni, 2015).

Compositionality beyond compounds

In the present article, we focused on compounds as compositional expressions that can be used to convey new meanings, and that reflect the cognitive operation of conceptual combination in the linguistic domain. However, compounds are only one instance of such expressions. We are confident that the general framework of our computationally-implemented, perceptually-grounded conceptual combination model can be adapted to other constructions related to the same phenomenon:

At the individual-word level, affixed words such as *rewrite* or *solidify* also are combinations of two meaning-bearing elements, the difference to compounds being that only one of the constituents (the stem) can occur as a free word, while the other (the affix) is a bound morpheme that cannot occur in isolation. Nevertheless, affixation is also a productive word-formation technique, and can thus be used to compositionally produce new meanings (such as *reinsult*). In empirical studies, it has been shown that a compositional distributional semantic model very similar to the CAOSS model (the FRACSS model, Marelli & Baroni, 2015) can successfully predict the meaning of such words, and empirical data such as explicit judgments and processing times. In this model, each affix is conceptualized as a separate weight matrix that is applied to the distributional vector representing the free word meaning of the stem (for a similar, linear-mapping based approach that does not require separate representations for each affix, see Baayen, Chuang, Shafaei-Bajestan, & Blevins, 2019). As long as vision-based representations for the stems are available, this framework can be adapted to model perceptually-grounded conceptual combination in affixed words, analogously to the CAOSS-based model introduced in the present study.

The same argument can be made for phrases consisting of more than one individual word, such as adjective-noun constructions. These expressions have been the focus of the earliest conceptual combination theories (Smith & Osherson, 1984), and arguably take an even more prominent position in this line of literature than compounds do (Murphy, 2002; Ran & Duimering, 2009). This is interestingly paralleled by the development of compositional distributional semantic models, which also started with a strong focus on adjective-noun combinations (Baroni & Zamparelli, 2010; Mitchell & Lapata, 2010) and have been demonstrated to predict explicit participant judgments on the sensibility of such phrases (Vecchi, Marelli, Zamparelli, & Baroni, 2017). Again, these models can be straightforwardly adapted to take vision-based representations as input.

Thus, since conceptual combination is defined as the combination of multiple meanings into a single new one, we are confident that variations of the model presented here can be adapted for any compositional process whose result can reasonably be described as a "single concept". At the level of

larger constructions, such as sentences or whole texts, psychological models of meaning construction usually shift towards a different level of description (mental models that include relations between many concepts, instead of conceptual combination; e.g. Johnson-Laird, 1983; Kintsch, 1988), and compositional distributional semantic models tend to describe very different phenomena with very different methods (e.g. Coecke, Sadzadeh, & Clark, 2010).

Conclusion

In the present study, we advocate the view that concepts are formed through all experience available with them (Zwaan & Madden, 2005; see also Günther et al., 2019). Some of their constituting and associated information might be linguistic, other information grounded in sensorimotor experience, with the relative proportion depending on the experience we make (Günther et al., 2018). When performing cognitive operations on these concepts, such as conceptual combination, we use the information and integrated experience available to us (Barsalou, 2008; Wu & Barsalou, 2009), as there is no reason to discard portions of it due to the modality through which it was acquired. In this perspective, processes involving the activation and mental manipulation of concepts – such as the processing of morphologically complex words – are routinely influenced by the available information from different sources. The present study empirically supports this view, by identifying effects of vision-based compositionality in four large-scale experiments employing three different, purely linguistic tasks. Due to the automatic nature of this concept-combination process, these perceptually-grounded compositional effects emerge even during the processing of existing, familiar complex words which already have a defined meaning in the reader's mind.

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., Marelli, M., & Crepaldi, D. (2015). The fruitless effort of growing a fruitless tree: Early morpho-orthographic and morpho-semantic effects in sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41, 1587–1596.
- Amenta, S., Marelli, M., & Sulpizio, S. (2017). From sound to meaning: Phonology-to-semantics mapping in visual word recognition. *Psychonomic Bulletin & Review*, 24, 887–893.
- Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, 116, 463–498.
- Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*, 2019.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12–28.
- 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.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical data base (CD-ROM)*. University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.
- Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445–459.
- 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., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of ACL 2014* (pp. 238–247). East Stroudsburg, PA: ACL.
- Baroni, M., & Zamparelli, R. (2010). Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 1183–1193). East Stroudsburg, PA: ACL.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 637–660.
- Barsalou, L. W. (2008). Grounding symbolic operations in the brain's modal systems. In G. R. Semin & E. R. Smith (Eds.), *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches* (pp. 9–42). New York, NY: Cambridge University Press.
- Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulations in conceptual processing. In M. D. Vega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 245–283). Oxford, UK: Oxford University Press.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- BNC Consortium. (2007). *The British National Corpus, version 3*. Oxford, UK: Bodleian Libraries. Retrieved from <http://www.natcorp.ox.ac.uk/>
- Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts. *Psychological Bulletin*, 143, 263–292.
- Bracci, S., Ritchie, J. B., Kalfas, I., & de Beeck, H. O. (2019). The ventral visual pathway represents animal appearance over animacy, unlike human behavior and deep neural networks. *Journal of Neuroscience*, 1714–18.
- Bruni, E., Tran, N.-K., & Baroni, M. (2014). Multimodal distributional semantics. *Journal of Artificial Intelligence Research*, 49, 1–47.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-

- quality, data? *Perspectives on Psychological Science*, 6, 3–5.
- Chamberlain, J. M., Gagné, C. L., Spalding, T. L., & Lõo, K. (2019). Detecting spelling errors in compound and pseudo-compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Advance online publication. doi: doi.org/10.1037/xlm0000748
- Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint; arXiv:1405.3531*.
- Coecke, B., Sadrzadeh, M., & Clark, S. (2010). Mathematical foundations for a compositional distributional model of meaning. *arXiv:1003.4394*.
- Connell, L., & Lynott, D. (2011). Interpretation and Representation: Testing the Embodied Conceptual Combination (ECCO) Theory. In A. Karmiloff-Smith, N. Nersessian, & B. Kokinov (Eds.), *Proceedings of the Third European Conference in Cognitive Science*. Sofia, Bulgaria: Cognitive Science Society.
- Connell, L., & Lynott, D. (2013). Flexible and fast: Linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & Review*, 20, 542–550.
- Costello, F. J., & Keane, M. T. (2000). Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science*, 24, 299–349.
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47, 1–12.
- de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, 48, 1–12.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255).
- de Saussure, F. (1916). *Course in general linguistics*. New York, NY: McGraw-Hill.
- Dima, C. (2015). Reverse-engineering Language: A Study on the Semantic Compositionality of German Compounds. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP2015)* (pp. 17–21). Lisbon, Portugal: ACL.
- 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.
- El-Bialy, R., Gagné, C. L., & Spalding, T. L. (2013). Processing of English compounds is sensitive to the constituents' semantic transparency. *The Mental Lexicon*, 8, 75–95.
- Estes, Z. (2003). Attributive and relational processes in nominal combination. *Journal of Memory and Language*, 48, 304–319.
- Feldman, L. B., Basnight-Brown, D. M., & Pastizzo, M. J. (2006). Semantic influences on morphological facilitation: Concrete-ness and family size. *The Mental Lexicon*, 1, 59–84.
- Firth, J. R. (1957). *Papers in linguistics, 1934–1951*. Oxford, UK: Oxford University Press.
- Fischer, M. H. (2012). A hierarchical view of grounded, embodied, and situated numerical cognition. *Cognitive Processing*, 13, 161–164. doi: 10.1007/s10339-012-0477-5
- Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association*, 105(2), 203–206.
- Franke, M. (2016). The evolution of compositionality in signaling games. *Journal of Logic, Language and Information*, 25, 355–377.
- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100, 25–50.
- Gagné, C. L. (2000). Relation-based versus property based combinations: A test of the CARIN theory and dual-process theory of conceptual combination. *Journal of Memory and Language*, 42, 365–389.
- Gagné, C. L. (2001). Relation and lexical priming during the interpretation of noun–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 236–254.
- 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, 71–87.
- Gagné, C. L., & Spalding, T. L. (2004). Effect of relation availability on the interpretation and access of familiar noun–noun compounds. *Brain and Language*, 90, 478–486.
- Gagné, C. L., Spalding, T. L., & Schmidtko, D. (2019). LADEC: The Large Database of English Compounds. *Behavior Research Methods*, 51, 2152–2179.
- Gelman, A., & Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *The American Statistician*, 60, 328–331.
- Glenberg, A. M. (2015). Few believe the world is flat: How embodiment is changing the scientific understanding of cognition. *Canadian Journal of Experimental Psychology*, 69, 165–171.
- Glenberg, A. M., & Robertson, D. A. (2000). Symbol Grounding and Meaning: A Comparison of High-Dimensional and Embodied Theories of Meaning. *Journal of Memory and Language*, 43, 379–401.
- Grice, P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics, 3: Speech Acts* (pp. 41–58). New York, NY: Academic Press.
- 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., Dudschig, C., & Kaup, B. (2018). Symbol grounding without direct experience: Do words inherit sensorimotor activation from purely linguistic context? *Cognitive Science*, 42, 336–374.
- 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). doi: 10.1371/journal.pone.0163200
- Günther, F., & Marelli, M. (2018). The language-invariant aspect of compounding: Predicting compound meanings across languages. In E. Cabrio, A. Mazzei, & F. Tamburini (Eds.), *Pro-*

- ceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-IT 2018) (pp. 230–234). Turin, Italy: Accademia University Press.
- Günther, F., & Marelli, M. (2019a). 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. (2019b, Jul). *A large dataset of timed sensibility judgments for existing and novel English compound words*. OSF. Retrieved from osf.io/7kynq doi: 10.17605/OSF.IO/7KYNQ
- Günther, F., & Marelli, M. (in press). Trying to make it work: Semantic effects in the processing of compound “nonwords”. *Quarterly Journal of Experimental Psychology*.
- Günther, F., Marelli, M., & Bölte, J. (in press). Semantic transparency effects in German compounds: A large dataset and multiple-task investigation. *Behavior Research Methods*.
- 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., Smolka, E., & Marelli, M. (2019). ‘Understanding’ differs between English and German: Capturing systematic language differences of complex words. *Cortex*, *116*, 168–175.
- Harris, Z. (1954). Distributional Structure. *Word*, *10*, 146–162.
- Hodgson, J. M. (1991). Informational constraints on pre-lexical priming. *Language and Cognitive Processes*, *6*, 169–205.
- Hollis, G. (2017). Estimating the average need of semantic knowledge from distributional semantic models. *Memory & Cognition*, *45*, 1350–1370.
- Jackendoff, R. (2002). *Foundations of knowledge*. Oxford, UK: Oxford University Press.
- James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance*, *104*, 130–136.
- Jenkins, J. J. (1954). Transitional organization: Association techniques. In C. E. Osgood & T. A. Sebeok (Eds.), *Psycholinguistics. A Survey of Theory and Research Problems* (p. 112–118). Bloomington, IN: Indiana University Press.
- Ji, H., Gagné, C. L., & Spalding, T. L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language*, *65*, 406–430.
- Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, MA: Harvard University Press.
- Jones, M. N., Kintsch, W., & Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, *55*, 534–552.
- 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.
- Juhász, 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.
- Kelter, S., & Kaup, B. (2012). Conceptual knowledge, categorization, and meaning. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning* (Vol. 3, pp. 2775–2805). Berlin, Germany: de Gruyter.
- Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*, 287–304.
- Kiela, D., Bulat, L., & Clark, S. (2015). Grounding semantics in olfactory perception. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)* (pp. 231–236). Beijing, China: ACL.
- Kiela, D., & Clark, S. (2015). Multi- and cross-modal semantics beyond vision: Grounding in auditory perception. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)* (pp. 2461–2470). Lisbon, Portugal: ACL.
- Kim, J. S., Elli, G., & Bedny, M. (2019). Furry hippos and scaly sharks: Knowledge of animal appearance among sighted and blind adults. *PsyArXiv preprint*. Retrieved from <https://doi.org/10.31234/osf.io/hw5pm>
- Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review*, *95*, 163–182.
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*, 10681–10686.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 1097–1105).
- Kroll, J. F., & Merves, J. S. (1986). Lexical Access for Concrete and Abstract words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *12*, 92–107.
- Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, *23*, 1089–1132.
- Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 876–895.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.
- 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.
- Lazaridou, A., Marelli, M., & Baroni, M. (2017). Multimodal word meaning induction from minimal exposure to natural text. *Cognitive Science*, *41*, 677–705.
- Lazaridou, A., Pham, N. T., & Baroni, M. (2015). Combining lan-

- guage and vision with a multimodal skip-gram model. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies* (pp. 153–163). East Stroudsburg, PA.
- Lebois, L. A., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? the context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science*, *39*, 1764–1801. doi: 10.1111/cogs.12174
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444.
- Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Rivista di Linguistica*, *20*(1), 1–31.
- Lenci, A. (2018). Distributional models of word meaning. *Annual Review of Linguistics*, *4*, 151–171.
- 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. (2014). The nature of compounds: A psychocentric perspective. *Cognitive Neuropsychology*, *31*, 8–25.
- Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language*, *84*, 50–64.
- Lopopolo, A., & Miltenburg, E. (2015). Sound-based distributional models. In *Proceedings of the 11th International Conference on Computational Semantics* (pp. 70–75).
- Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*, 273–302.
- Lucas, M. (2000). Semantic priming without association. *Psychonomic Bulletin & Review*, *7*, 618–630.
- Lynott, D., & Connell, L. (2010). Embodied conceptual combination. *Frontiers in Psychology*, *1*, 212.
- 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., 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.
- 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.
- Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, *101*, 3–33.
- Martin, D. I., & Berry, M. W. (2007). Mathematical Foundations Behind Latent Semantic Analysis. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 35–56). Mahwah, NJ: Erlbaum.
- McNamara, D. S., Cai, Z., & Louwerse, M. M. (2007). Optimizing lsa measures of cohesion. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 379–400). Mahwah, NJ: Erlbaum.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv:1301.3781v3*.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). 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.
- Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., & Baayen, R. H. (2017). Discrimination in lexical decision. *PLoS ONE*, *12*(2), e0171935.
- 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. (1990). Noun phrase interpretation and conceptual combination. *Journal of Memory and Language*, *29*, 259–288.
- Murphy, G. L. (2002). Conceptual Combination. In G. L. Murphy (Ed.), *The Big Book of Concepts* (pp. 443–475). Cambridge, MA: MIT Press.
- Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling. *Ecography*, *37*, 191–203. doi: 10.1111/j.1600-0587.2013.00205.x
- Ostarek, M., & Huettig, F. (2019). Six challenges for embodiment research. *Current Directions in Psychological Science*, *0963721419866441*.
- Paivio, A. (1966). Latency of verbal associations and imagery to noun stimuli as a function of abstractness and generality. *Canadian Journal of Psychology*, *20*, 378–387.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, UK: Oxford University Press.
- 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.
- Petilli, M. A., Günther, F., Vergallito, A., Ciaparelli, M., & Marelli, M. (2019). Data-driven computational models reveal perceptual simulation in word comprehension. *PsyArXiv preprint*. Retrieved from <https://doi.org/10.31234/osf.io/98z72>
- Pezzelle, S., Shekhar, R., & Bernardi, R. (2016). Building a bag-pipe with a bag and a pipe: Exploring conceptual combination in vision. In *Proceedings of the 5th Workshop on Vision and Language* (pp. 60–64).
- Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., ... O'Toole, A. J. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences*, *115*, 6171–6176.
- Ramscar, M., Hendrix, P., Love, B., & Baayen, R. (2013). Learning is not decline: The mental lexicon as a window into cognition across the lifespan. *The Mental Lexicon*, *8*, 450–481.
- Ran, B., & Duimering, P. R. (2009). Conceptual Combination: Models, Theories, and Controversies. In S. P. Weingarten &

- H. O. Penat (Eds.), *Cognitive Psychology Research Developments* (pp. 39–64). New York, NY: Nova Science.
- Rastle, K. (2018). The place of morphology in learning to read in English. *Cortex*, *116*, 45–54.
- Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, *23*, 942–971.
- Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes*, *15*, 507–537.
- Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*, 1090–1098.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*, 510–532.
- 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.
- Schmidtke, D., & Kuperman, V. (2019). A paradox of apparent brainless behavior: The time-course of compound word recognition. *Cortex*, *116*, 250–267.
- Schmidtke, D., Matsuki, K., & Kuperman, V. (2017). Surviving blind decomposition: A distributional analysis of the time-course of complex word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Advance online publication. doi: 10.1037/xlm0000411
- 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.
- Schreuder, R., & Baayen, R. H. (1995). *Modeling morphological processing*. Hillsdale, NJ: Erlbaum.
- Smith, E. E., & Osherson, D. N. (1984). Conceptual Combination with Prototype Concepts. *Cognitive Science*, *8*, 337–361.
- Smolka, E., Preller, K. H., & Eulitz, C. (2014). 'Verstehen' ('understand') primes 'stehen' ('stand'): Morphological structure overrides semantic compositionality in the lexical representation of German complex verbs. *Journal of Memory and Language*, *72*, 16–36.
- Spalding, T. L., Gagné, C. L., Mullaly, A. C., & Ji, H. (2010). Relation-based interpretation of noun-noun phrases: A new theoretical approach. *Linguistische Berichte Sonderheft*, *17*, 283–315.
- Striem-Amit, E., Wang, X., Bi, Y., & Caramazza, A. (2018). Neural representation of visual concepts in people born blind. *Nature Communications*, *9*, 5250.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.
- Thagard, P. (1984). Conceptual combination and scientific discovery. In P. Asquith & P. Kitcher (Eds.), *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* (Vol. 1, pp. 3–12). East Lansing, MI: Philosophy of Science Association.
- Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188.
- Vecchi, E. M., Marelli, M., Zamparelli, R., & Baroni, M. (2017). Spicy adjectives and nominal donkeys: Capturing semantic deviance using compositionality in distributional spaces. *Cognitive Science*, *41*, 102–136.
- Vedaldi, A., & Lenc, K. (2015). Matconvnet: Convolutional neural networks for Matlab. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 689–692).
- Westbury, C. (2016). Pay no attention to that man behind the curtain: Explaining semantics without semantics. *The Mental Lexicon*, *11*, 350–374.
- 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.
- Wisniewski, E. J., & Love, B. C. (1998). Relations versus properties in conceptual combination. *Journal of Memory and Language*, *38*, 177–202.
- Wu, L.-L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, *132*, 173–189.
- Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European conference on computer vision* (pp. 818–833).
- Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 586–595).
- Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of action and perception in memory, language, and thinking* (pp. 224–245). Cambridge, UK: Cambridge University Press.
- Zwitserslood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes*, *9*, 341–368.

Supplementary Material A: Follow-Up Analyses

Follow-Up Analysis 1: Conformity with Previous Results

Lexical decision data obtained from the English Lexicon Project (Balota et al., 2007) – albeit a considerably larger dataset, since it was not required that vision-based representations were available – was already employed in the study by Günther and Marelli (2019a), which focused on investigating language-based measures of compositionality. While the study by Günther and Marelli (2019a) observed effects of modifier and head composition, the same pattern did not emerge in the present study (see Table 4 of the main article). However, these differences can potentially be attributed to two different factors: On the one hand, there are slight differences concerning the parametrization of the language-based model between the present study and the previous one. The CAOSS model in the previous study was estimated from a smaller training set, and other frequency-cutoff criteria were applied for the training of word embeddings. On the other hand, the present study employs a systematically selected item set that is significantly smaller than the previous one, since only compounds with visual representations associated to both their constituents have been investigated here.

To ensure that the parameter choices for the language-based model did not cause changes in the pattern of results, we replicated the analysis by Günther and Marelli (2019a) with our present language-based model. We first estimated a Linear Mixed Effects Model to predict logarithmic lexical decision response times for the larger set of all 1,443 compounds included in all three behavioral datasets which occurred more than 50 times in the source corpus (i.e., also considering compounds for which vision-based representations were not available). As predictors, this model contained random intercepts for modifiers and heads, as well as fixed effect parameters for compound length, modifier, head, and compound frequencies, and (language-based) modifier and head composition (Günther & Marelli, 2019a). As in the previous study, both constituent-composition parameters significantly predicted response times (see Table 8, middle column). Thus, differences in the model setup are not responsible for the absence of modifier and head composition effects in the present study.

In a next step, we extended this analysis to the two additional experimental paradigms that were not previously tested by Günther and Marelli (2019a). In the timed sensibility judgment task, we again observed significant effects for both modifier and head composition (Table 8, right column). In the naming task however, we observed no such effects (Table 8, left column).

We next examined whether the differences in results between the present study and the one by Günther and Marelli (2019a) can be traced back to the fact that the present study investigated a smaller item set. To examine this pos-

sibility, we first estimated the same models as described in the previous step on the subset of compounds for which all vision-based measures are available. In the timed sensibility judgment task, both effects were still significant (modifier composition: $t = -2.14$, $p = .033$; head composition: $t = -2.45$, $p = .015$). However, in the lexical decision task, the constituent composition effects were no longer significant (modifier composition: $t = -1.52$, $p = .130$; head composition: $t = -0.64$, $p = .524$). The same was still true for the naming task (modifier composition: $t = -1.28$, $p = .203$; head composition: $t = -0.53$, $p = .597$).

Thus, modifier and head composition effects in lexical decision are not significant for this subset. However, this can either be caused by a lack of statistical power due to the smaller size of the item set, or by a systematic difference between the subsets. We therefore tested whether the effects of language-based modifier and head composition differ between the two subsets of compounds (Gelman & Stern, 2006): The one for which all vision-based measures are available, and the one for which they are not. We performed two separate likelihood-ratio tests for each of the models reported in Table 8, comparing the respective models (each of which now also included a dummy variable encoding whether the vision-based measures are available for the compound) to (a) a model that additionally contained an interaction between modifier composition and this dummy variable, and (b) a model that additionally contained an interaction between head composition and this dummy variable. None of those tests was significant (all $\chi^2(1) < 0.62$, all $p > .429$). This suggests that the absence of language-based constituent composition effects in the lexical decision tasks in the subset can be attributed to a lack of statistical power, rather than systematic differences between the subsets.

However, these analyses yielded another, interesting result: The main effect for the dummy variable itself was significant for the lexical decision task ($b = -0.02$, $t = -2.47$, $p = .014$) and closely failed to reach significance for the timed sensibility task ($b = -0.02$, $t = -1.89$, $p = .059$), indicating faster lexical decisions for compounds for which the vision-based measures are available (which are exactly those compounds for which both constituents have a vision-based representation). Including this dummy variable did not affect the significance level of the other parameters in the model.

Follow-Up Analysis 2: Imagery Effects

We followed up on this finding by subjecting it to a more rigorous test. For each task, we estimated a baseline mixed-effect model, predicting logarithmic reaction times from all language-based predictors: length, all frequencies and family sizes, and all language-based semantic measures in Table 2 of the main article, as well as random intercepts for the modifiers and heads. We then compared, for each task, this baseline model to a model that additionally contained the

Table 8

Model parameters – regression weights (t-values) – predicting logarithmic response times in pre-analysis 1. Only significant parameters are displayed ($p < .05$).

parameter	naming		lexical decision		timed sensibility	
intercept	6.76	(170.94)	7.03	(146.77)	7.08	(120.54)
length	0.03	(13.76)	0.02	(9.85)	0.02	(5.30)
compound frequency	-0.02	(-12.84)	-0.04	(-18.26)	-0.04	(-16.33)
modifier frequency	-0.01	(-6.79)	-0.01	(-5.51)	-0.01	(-2.29)
head frequency	-0.01	(-4.76)	0.01	(-3.11)	0.01	(2.23)
modifier composition			-0.06	(-2.09)	-0.10	(-2.36)
head composition			-0.07	(-2.04)	-0.13	(-3.18)

dummy variable encoding whether all vision-based measures are available for the compound in a likelihood-ratio test.

For the naming task, including the dummy variable did not improve the model fit ($\chi^2(1) = 0.08$, $p = .782$). For the lexical decision task, including the dummy variable did significantly improve the model fit ($\chi^2(1) = 5.78$, $p = .016$), but it failed to reach significance for the timed sensibility judgment task ($\chi^2(1) = 3.59$, $p = .058$). There was no additional processing advantage for the 384 items for which also a compound visual representations was available ($\chi^2(1) = 0.89$, $p = .345$ for timed sensibility judgments, ($\chi^2(1) = 0.001$, $p = .967$ for lexical decisions).

Thus, at least in the lexical decision task, just having available perceptual information for the constituents speeds up processing. This finding is in line with a large body of literature on concreteness effects (e.g. Kroll & Merves, 1986), which are usually explained in terms of an imagery ef-

fect – words with associated verbal *and* perceptual information (i.e., for which a mental image can be created) are processed faster than words for which only verbal information is available (Paivio, 1966, 1986). While previous studies have demonstrated processing advantages for imageable complex words (Feldman et al., 2006), our results extend upon previous findings by shifting the locus of the effect to the sub-lexical level: Complex words are processed faster if mental images can be created for their constituents, indicating sub-lexical semantic influences on processing (in line with previous studies observing semantic effects of the constituents, such as stem valence effects; Schmidtke, Matsuki, & Kuperman, 2017; Schmidtke & Kuperman, 2019). Of course, imageable constituents are also extremely likely to result in an imageable combined representation, but we find no evidence for an additional benefit at the compound level.