Distributional Semantics meets Embodied Cognition: Flickr® as a database of semantic features

Selected Papers from the 4th UK Cognitive Linguistics Conference
http://uk-cla.org.uk/proceedings
Pages: 18 – 35

MARIANNA BOLOGNESI

International Center for Intercultural Exchange
marianna.bolognesi@gmail.com

Abstract

Distributional models such as Latent Semantic Analysis (LSA, Landauer, Dumais 1997) generate semantic spaces based on the co-occurrences of words in linguistic contexts. The semantic representations that emerge from these models are based solely on linguistic information, leaving aside the information that we retrieve from perceptual experiences. The proposed analytical approach applies the methods of distributional semantics to Flickr®, a corpus of images enhanced with metadata (tags), expressing a wide range of concepts, including perceptual features triggered by the experiences captured in the photographs. A case study on the domain of colors shows how a distributional analysis based on Flickr® can produce semantic representations for color terms that better resemble the similarity judgments provided by humans, when compared to those that emerge from distributional models based on solely linguistic information.

Key words: distributional semantics, grounded cognition, corpus analysis, annotated images.

1. Introduction

In the last decades, symbolic and disembodied accounts of cognition (e.g. Fodor 1975; Haugesland 1985; Pylyshyn 1984) have been challenged by embodied theories, which support with overwhelming experimental evidence the claim that knowledge is grounded in the brain's modal system, and also that language comprehension requires, to a certain extent, the activation of semantic information stored during our bodily experiences in the world (e.g., Barsalou et al. 2008; Glenberg 1997; Pecher et al. 2003; Vigliocco et al. 2009; Zwaan 2004). The

scientific shift towards embodied positions, as opposed to symbolic ones, has put the supporters of the Distributional Hypothesis in a critical situation.

The Distributional Hypothesis states that the meaning of a word can be derived by looking at the linguistic contexts in which the word occurs (Harris 1954; Firth 1957; Miller and Charles 1991). It follows that the degree of semantic relatedness between two words may be inferred by looking at the number of contexts that the two words share. For example, if we observe the linguistic contexts in which the words *book*, *manual*, and *umbrella* usually appear, we note that *book* and *manual* share more contexts than *book* and *umbrella* or *manual* and *umbrella*. Thus, there is a greater similarity between *book* and *manual*, as opposed to *book* and *umbrella* or *manual* and *umbrella*. The Distributional Hypothesis is implemented by the methods of distributional semantics, where meanings come to be defined by vectors that keep track of the co-occurrences of words in large collections of texts.

Although distributional models can reproduce the human behavior in specific language-related tasks with a high degree of fidelity, it has been argued that the Distributional Hypothesis, as we know it, does not constitute a valid theory of meaning because it models semantic representations starting from words' occurrence with other words, and this mechanism establishes a vicious circle that does not explain how the linguistic symbols are connected to the world (*symbol-grounding problem*: Harnad 1990; Glenberg and Robertson 2000). However, a new generation of distributional models has recently prompted alternative solutions to overcome the *symbol-grounding problem*, proposing hybrid models based on integrated analyses of semantic information retrieved from language and images (e.g. Feng and Lapata 2010; Bruni et al. 2011), under the assumption that images and their related visual features are a valid proxy for actual perceptual experience. Nevertheless, the integration of visual features retrieved from images can be controversial for two main reasons. Firstly, our conceptual system does not act like a camera, recording holistic images, but it rather selects and interprets specific aspects of each experience (e.g. Barsalou 2012). The second controversial issue is that reducing all non-verbal information that is missing in the classic distributional models to visual components of an image does not provide a trustworthy take on how we really process and represent meaning.

In this contribution, I propose the use of a new source of data for distributional analyses: I argue that the components of experience that populate our conceptual system and allow us to interpret and make sense of the world are fairly well approximated by the speaker-generated tags that users attribute (during the *tagging* process) to their photographs uploaded on Flickr®[i], the social network for images and photo/video hosting service powered by Yahoo!.

This hypothesis is tested through the implementation of a specific distributional model where concepts[ii] cluster according to their similarity, computed in terms of the covariance of tags across thousands of images. The creation of this model will be explained and exemplified through a case study based on color terms. The method used for its implementation is based on the Distributional Hypothesis and it is framed within the family of methods implemented in distributional semantics. However, a full explanation of this method and a deeper evaluation against other state-of-the-art distributional methods still needs to be provided. Here, the results of this first explorative case study based on color terms will be reported, and compared to the results of the same study conducted on two traditional distributional models that are based on exclusively linguistic data. In this way, the semantic representations of color terms that emerge from the corpus of annotated images, and therefore encompass perceptually-derived information, will be compared to the semantic representations that emerge from our linguistic use of color terms. The two linguistic distributional models, which will act as control for the outputs obtained with the model based on Flickr®, are: (i) an unstructured model that retrieves word occurrences in terms of manifestation of a word in a document (LSA, Landauer and Dumais 1997), and (ii) a structured model, which takes into account the syntactic patterns and the semantic collocates that surround a word's occurrence (DM, Baroni and Lenci 2010). Observing the three types of semantic representations makes it possible to compare the conceptual relatedness that emerges between two words from: (i) their simple linguistic co-occurrence, (ii) their occurrence in syntactic patterns with other semantic collocates, (iii) their more general perceptual experience, reported through images. This will provide empirical data for a discussion of the symbol-grounding problem and its implications on the nature of semantic representations.

2. Method

2.1 Flickr®: a database of semantic information and repository of social cognition

Founded in 2004, and still growing fast[iii], Flickr® is a social network for images, where users can upload personal photographs and enhance them with metadata such as titles, captions, and tags. Such information facilitates the organization of this data into virtual albums and their retrieval by other users. Flickr® tags are short textual labels (one-word or compounds), expressed in any supported language (Unicode, UTF-8). Each photograph can be enriched with a maximum of 75 tags, although the average number of tags attributed to each image is much lower.

Having been created by users, rather than by professional annotators, the tags in Flickr® constitute a social phenomenon whose analysis can

lead to the creation of tools for social knowledge representation, and shed light on the structure of human distributed cognition.

A Flickr® tag is a lexical label used for expressing a correspondent concept. Within a specific photograph, which is to say within an individual experience encoded in a photograph, a tag constitutes a manifestation of a concept (token), which contributes to the description of a portion or an aspect of the experience represented by the photograph.

According to grounded accounts of cognition, meanings are rooted in experience, and grounded in our sensorimotor system, which makes sense of the world through our bodies. Merging this assumption to the Distributional Hypothesis, it can be argued that, given an experiential context, each component on which our attention focuses is a concept that contributes to define a salient aspect of that situation[iv]. Because the whole situation can be represented as a set of salient features (or salient concepts), then, within an individual situation, each salient concept defines and comes to be defined (at least partially) by the other salient concepts that are experienced contextually to that situation. *Tout se tient*. Everything is tied together. But, also, everything is rooted in the perceptual experience.

For example, the concept of *bread*, expressed in Flickr® through the tag "bread", might be fairly well defined by a large enough sample of pictures that includes this tag. Operationalizing this idea on the basis of the Distributional Hypothesis, it can be argued that the concept of *bread* is fairly well approximated by the (weighted) myriad of tags that appear together with the tag "bread", across the Flickr® photographs (e.g. "food", "loaf", "baking", "breakfast", "cheese", "yummy"). This is the principle that underlies the implementation of distributional semantic spaces, or vector spaces (see Sahlgren 2006 for a review), here applied to Flickr®. When we think about *bread*, we immediately activate associated properties and features (which are other concepts), and this spread of activation arguably involves entities, events, emotions, and locations that have been perceived contextually with *bread* in previous experiences. In this perspective, the interpretative process that allows us to understand and make sense of the world is achieved thanks to selective attention focused on specific features of an episode (here represented by tags), which are then integrated with other semantic information retrieved from previous experiences.

From this perspective, our memory is a repository of concepts (now intended as types, rather than individual manifestations of tokens), where each concept is (at least partly) defined by the concepts that are perceived together with it in the myriad of situations that we experience.

2.2 Implementing a distributional model based on Flickr® tags.

The amount of labeled image collections available online is constantly increasing. The semantic information provided by humans gathered in these databases is fundamental for the implementation of robust tools in computer vision research. Such databases are built with the purpose of stimulating coordinated descriptions of the same image between two or more users (e.g. ESP game, Google Image Labeler, LabelMe[v]). In fact, some of these tools appear as real time games, implemented for gathering spontaneous annotations without imposing a strict experimental setting on the participants. However, the purpose of such activities, and the instructions provided, can arguably constrain the range of free associations that a user might attribute to a given visual stimulus. Flickr® was not built with this purpose, and the tags that are attributed to each image include a much wider range of concepts. Moreover, while, in the image collections mentioned above, the visual stimuli are randomly selected, in Flickr®, the uploaded images are the user's personal experiences reported on the social network in the form of a photograph. Each image in Flickr® is a visual proxy for the actual photographer's experience. As a consequence, the tags attributed to these images reflect the set of semantic features[vi] perceived during the actual experience with an arguably higher degree of reliability.

Although there is still no proper classification and standardization of all the types of tags that users attribute to photographs in Flickr®, from a general perspective, it can be stated that the encoded semantic information includes visually-related features (settings, objects, people, actions, events, properties), as well as mental states, emotions, and other features expressing concrete and abstract concepts that are not directly captured by the image, but are triggered during the conceptual processing of the actual experience that lies behind the photograph. This type of information makes Flickr® tagsets[vii] richer and more variegated than the keywords that are attributed to random photographs in other databases of annotated images. For example, in Flickr®, a picture can be tagged with the word "California" and represent a dog photographed in a living room of a friend's house in Los Angeles (see Steels 2006). This phenomenon, particularly common for tags that express geographical locations (*geotags*), supports the idea that tagging Flickr® photographs is a process that triggers the re-activation of the real experience. Thus, arguably, Flickr® tags include concepts that are derived from bottom-up perceptual processing, thanks to which we tag perceptual features in an image, and tags that express concepts derived from top-down conceptual processing, thanks to which we integrate perceptually-derived information with other semantic information retrieved from memory. Thus, tagsets represent information beyond a description of the simplistic visual features, which indicates that Flickr® users elaborate the experience encoded in the photograph on a deep conceptual level, rather than just pointing out visual features presented in the image.

As they are unsupervised and unfiltered, the metadata used for image annotation in Flickr® are much more complex to analyze than the instructed and clean sets of words used for annotating images in the databases described above. But the amount and the variety of information contained in Flickr® is much larger than the information contained in other databases, and, in a sense, more spontaneous and realistic, and more likely to resemble portions of the structure of our distributed cognition.

Unlike other social networks, since the beginning of its virtual life, Flickr® published a set of freely available APIs (Application Programming Interfaces). These tools allow external programmers to implement applications and access the database's content in order to post it in multiple locations of the web, as long as the privacy restrictions and the copyrights are respected, and, in some cases if the calls are signed with a valid API key.

The theoretical assumptions and the availability of this rich lexical and conceptual resource make Flickr® an optimal database of perceptual and conceptual information for distributional analyses. The case study described below suggests its reliability, showing how the inherently perceptual domain[viii] of colors (primary and secondary) is organized in a structure that reflects the distribution in Newton's color wheel (or in a rainbow). Such structure emerged automatically from the distributional analysis of color terms based on Flickr® tags.

2.3 The creation of the sub-corpus

Flickr® contains billions of photographs and this number increases every day. Therefore, as a first step for conducting a distributional analysis, it was necessary to sample the database and download a corpus of tagsets through the API Flickr.Photo.Search.

As explained in its documentation, this method returns the metadata associated to the photos that match the established criteria (e.g. date of upload, tags included, the number of photographs retrieved etc.) in different formats (e.g. XML tree). The data retrieval can be facilitated by an open source command-line utility that has been implemented by Buratti for unattended downloads of metadata from Flickr.com. This powerful tool is hosted on code.google.com[ix].

For the case study on the domain of colors, a million tagsets were downloaded for each of the six color terms analyzed (*red*, *orange*, *yellow*, *green*, *blue*, *purple*), which means that the corpus for the

analysis consisted of 6 million photos (or better 6 million tagsets associated to as many photos), where each color term appeared a million times. Tagsets featuring more than one color were excluded from the distributional analysis, in order to avoid the vicious circle of having, for example, the tag *red* as a semantic dimension for the meaning of *orange*, and, at the same time, the tag *orange* as a dimension for defining the meaning of *red*.

The downloaded tagsets belonged to photographs that were uploaded on Flickr® between January 2007 and December 2011. Around 17,000 tagsets were downloaded for each of the 12 months of each of the 5 years taken into account (the exact amount varied, depending on the availability of photographs uploaded each month).

The corpus was analyzed with *R,* an open source software package for statistical analyses. As a first step, each tagset was cut at the fifteenth tag: only the first 15 tags associated to each photograph were considered salient features for describing the represented episode. As a result, each tagset consisted of a number of tags ranging between 1 and 15.

Secondly, the redundant tagsets were eliminated. This means that only unique tagsets were kept. This step was performed because, in Flickr®, it is common to see tagsets copied and attached to entire batches of pictures taken by the same user. The noise that such redundancies bring into the data analysis is comparable to the same phenomenon that provokes unexpected outcomes while dealing with web-based linguistic corpora. In order to avoid biased frequency values determined by this phenomenon, only the unique tagsets were kept.

At this point, the distribution of the number of tags attributed to each picture would have been positively skewed if the maximum number of tags (75 tags per photo) was kept. However, by cutting the tail after the fifteenth tag, the distribution of the number of tags associated to each photograph peaked around the value 8, which means that most of the pictures had 8 tags (Mode=8, Mean=7.8, SD=3.1).

Then, a second type of filter was applied to the ranking of the color terms: only those tagsets where the color term was listed among the first 5 tags were kept. This step was performed after having observed that, in some tagsets, the color terms (the target of the analysis) appeared very late in the tagging process. It was assumed that for those images, the tag expressing the color was not a very salient feature in the user's mind, and therefore such image was not representative for the concept of that color.

After these operations, the resulting database consisted of 885,242

unique tagsets of fifteen (or less) tags each, showing one of the six colors among the first five tags. An excerpt of the resulting database is reported in Figure 1.

Figure 1: Excerpt of the database of tagsets showing the tag "red", used for the distributional analysis.

2.4 Contingency table

The collected data was organized in a contingency table, whereby the 6 color terms were listed in columns, and all the other unique tags used in the 885,242 tagsets were listed in rows (an excerpt is reported in Figure 2).

The list of unique tags included 473,490 items. Of these, 256,160 tags appeared only once or twice. Because most of these tags were actually non-words, or words spelled incorrectly, these tags were not considered significant, and were excluded from the contingency table. The resulting table was a 217,330 by 6 matrix (excluding headers). In the cells, the value of the Square Pointwise Mutual Information (SPMI, Bouma 2009) was computed, normalized by multiplying the squared joint frequency for the sample size (N). The obtained value approximates the likelihood of finding a color term and each of the tags listed on the rows appearing together in a tagset, taking into account the overall frequency of each of the two tags in the corpus, the frequency of their co-appearance within the same tagsets, and the corpus size. Negative values were raised to zero, a common practice in distributional semantics (e.g. Baroni, Lenci 2010).

At this point, each color term was defined by a list of coordinates (a column of the table), where each coordinate indicated the color's relation with each co-occurring tag in the corpus of tagsets.

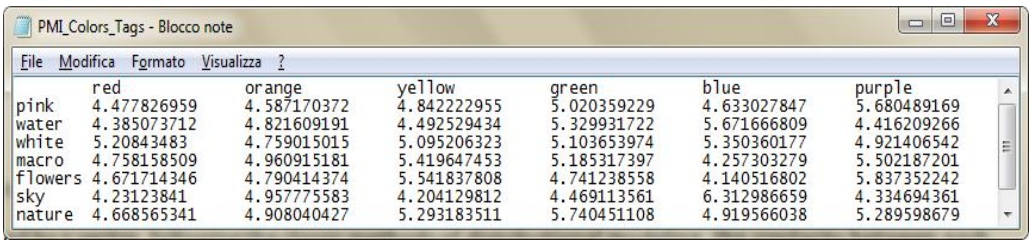


Figure 2: Excerpt from the contingency table showing the normalized SPMI value between each of the color terms and each of the other tags appearing across the downloaded tagsets.

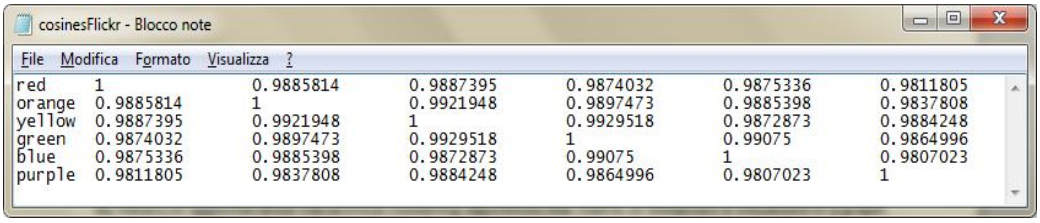
Sorting the values from the higher to the lower for each color terms allowed us to observe which concepts were most commonly associated to each color term. This operation also highlighted the common problems that arise when dealing with a web-based corpus that was not built for the purpose of collecting semantic features. The first of these problems is the fact that Flickr® tagsets often include tags expressed in more than one language. As a consequence, among the tags that most frequently appeared together with the six colors analyzed, we found their equivalents in other languages (e.g. the tags “rojo”, “rouge”, “rosso”, “rot”, and “vermelho”, appearing in the same pictures that were tagged with “red”). Another bias that was observed by looking at the tags that most frequently appear together with each color tag, and that can arguably generalized to other distributional analyses based on Flickr®, was the presence of keywords denoting the brand of the camera or technical details related to the photograph, such as “Nikon”, “Canon” or “macro”. These meta-tags lie outside the purpose of the analyses and therefore were manually removed. Other types of biased data were removed by simply cutting the lower tail of the ordered list of co-occurring tags for each color term. However, after this step, the contingency table had to be re-computed. The biases that were avoided by removing the low ranked tags include: tags with spelling mistakes (e.g. “lavander” instead of “lavender”), complex concepts expressed in a single tag without segmenting the lexical units (e.g. “firstdayofschool”), and proper names or nicknames (e.g. “Mike”, “Ginny”).

After that, the lists of co-occurring tags were ordered from higher to the lower SPMI value, with the top 100 co-occurring tags for each color term manually classified according to a simple ad-hoc created taxonomy. It appeared that the top 100 concepts that most commonly appear together with each color term across Flickr® can be grouped into seven categories: natural entities (such as “flower”, “tree”, “water”, “sun”, “ocean”, “garden”, “forest”, “landscape”, “cloud”); other colors (such as “pink”, “white”, “black”, “brown”); periods of time (such as “sunset”, “winter”, “night”); linguistic compounds (such as “sox” for “red”; “county” for “orange”), other concrete entities (such as “car”, “girl”, “city”, “eyes”), events (“wedding”, “Christmas”) and other modifiers (such as “beautiful”, “old”, “sexy”).

2.5 Calculating similarities

Each column of the updated contingency table at this point was a list of values that defined the behavior of a color term against each of the co-occurring tags across Flickr. In mathematical terms, each list of values (i.e. each column in the contingency table) is a vector. Following the practice of distributional modeling, the similarity between each two concepts was interpreted as the geometrical proximity between the two concepts' vectors. This was operationalized through the computation of the cosine, a coefficient that ranges between 0 and 1: the higher the cosine, the closer the two vectors are to one another, and the more the two concepts are similar.

Figure 3 reports the table of cosines between each two colors' vectors:


cosinesFlickr - Blocco note

File Modifica Formato Visualizza ?

red	1	0.9885814	0.9887395	0.9874032	0.9875336	0.9811805
orange	0.9885814	1	0.9921948	0.9897473	0.9885398	0.9837808
yellow	0.9887395	0.9921948	1	0.9929518	0.9872873	0.9884248
green	0.9874032	0.9897473	0.9929518	1	0.99075	0.9864996
blue	0.9875336	0.9885398	0.9872873	0.99075	1	0.9807023
purple	0.9811805	0.9837808	0.9884248	0.9864996	0.9807023	1

Figure 3: Cosines between each two color terms. The matrix is symmetrical. On the diagonal, there are the highest values showing the geometrical proximity between each color vector and itself (equal to 1).

By means of agglomerative hierarchical clustering algorithms, the matrix of distances was analyzed and the outcomes were visualized in the graph reported in Figure 4. The length of the arches that connect each two colors reflects the geometrical proximity between the two vectors: the shorter the arch between two words, the closer the two vectors, and therefore the greater the distributional similarity between the two concepts.

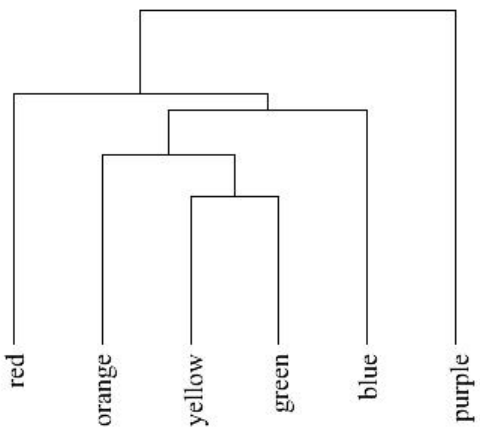


Figure 4: Graph representing the table of cosines by means of hierarchical clustering.

The distribution of the colors that emerged from the analysis based on Flickr® tagsets resembled the distribution of the wavelengths perceived by the three types of cones that characterize the human eye, which make us sensitive to three different spectra (Figure 5). In other words, the conceptual relatedness among colors emerging from the analysis based on annotated images uploaded on Flickr® resembled the actual proximities (expressed in nanometers) among the wavelengths of the spectrum of light that is visible to humans, thanks to which we perceive colors. Furthermore, in the scientific literature on color representations, it has been shown that the similarity judgments among color terms obtained from sighted participants yield to the same distribution, also represented in Newton's color circle (e.g. Izmailov and Sokolov 1992; Shepard and Cooper 1992).

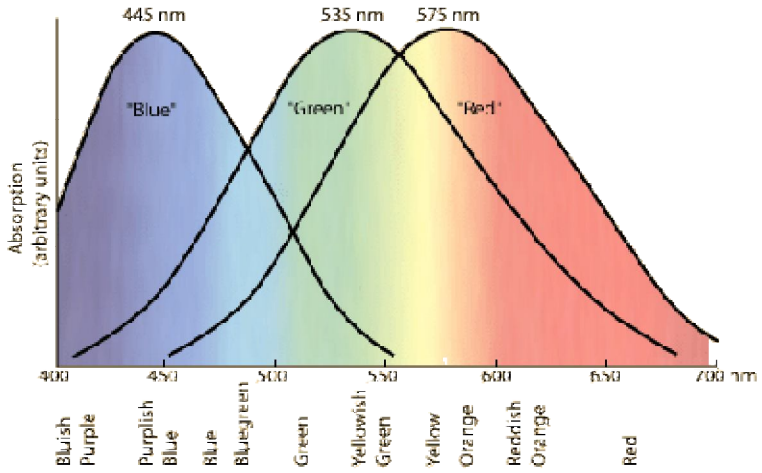


Figure 5: the color spectrum.

This resemblance suggests that applying the Distributional Hypothesis to the semantic information encoded in Flickr® tagsets might be a viable route for modeling the cognitive processes of how humans store and organize the semantic information retrieved from experience. The case study showed that the distances between colors emerging from the analysis of color terms expressed through tags resemble the distances between the light waves that our eyes perceive when we experience colors, and the similarity judgments that we attribute to colors when these are presented as word pairs. The next step was to test whether such perceptual information is encoded by language and can be triggered by existing distributional models based on corpora of texts, rather than images.

3. Colors in language: LSA and DM

The cosines between vectors indicating color terms, obtained with Latent Semantic Analysis (see Landauer, Dumais 1997 for a detailed

documentation on this method), are reported in Figure 6, together with the graph that shows the distributional similarities by means of hierarchical clustering algorithms. In Figure 7, instead, are the reported cosines and graph of the distributional analysis conducted with Distributional Memory (see Baroni, Lenci 2010 for detailed documentation on this method).

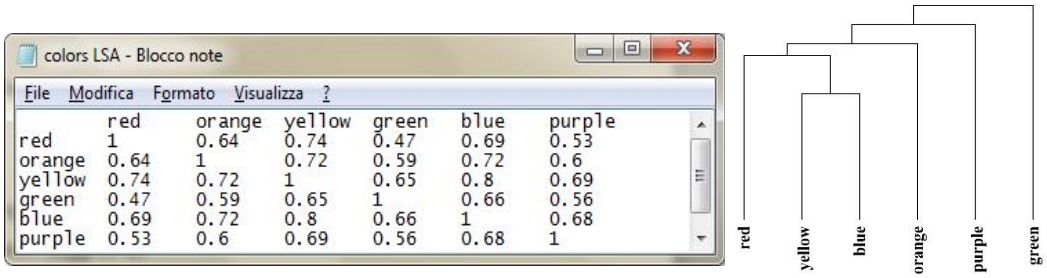


Figure 6: Similarities (cosines) between color terms, emerging from a LSA distributional analysis.

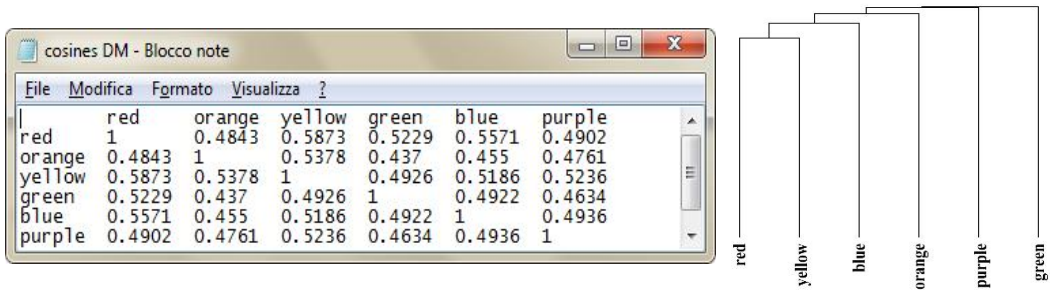


Figure 7: Similarities (cosines) between color terms, emerging from a DM distributional analysis.

As indicated by Figures 6 and 7, the semantic information encoded in language and investigated with distributional methods can be visualized in word spaces that do not seem to resemble the similarities retrieved from visual contexts. This seems to be true for both distributional models that retrieve semantic information from 'bags of words' (LSA, Fig. 6), as well as for distributional models that retrieve semantic information from a combination of syntactic patterns and semantic collocates (DM, Fig. 7). On the other hand, the two models based on linguistic corpora produce fairly similar word spaces, where the three primary colors (*red*, *yellow*, and *blue*) seem to cluster together and are followed by the three secondary colors. In both models, *green* is the farthest color, which means that is the least similar to the others in terms of linguistic distributions.

3.1 Correlation coefficients between representations

As expected from the observation of the graphs, the semantic representations of color terms that emerged from the two distributional models based on linguistic corpora (LSA and DM) had a strong and positive degree of correlation ($r=.88$, $n= 6$, $p=.02$). However, calculating the correlation between the tables of similarities, we could also observe strong and positive degrees of correlation between the semantic representations emerging from Flickr® tagsets and, respectively, LSA ($r=.80$, $n=6$, $p=.05$) and DM ($r=.83$, $n=6$, $p=.04$).

The very high coefficients of correlation between LSA and DM support the overall consistency and robustness of these two models: although they are based on different linguistic corpora and different techniques, which allow the retrieval and analysis of different linguistic contexts, these two models provide very similar outcomes. Moreover, the fact that the tables of similarities obtained with LSA and DM show fairly high degrees of correlation with the one obtained from Flickr® suggests that language by itself indeed provides a rich source of information that fairly accurately matches the perceptual information triggered by the visual contexts gathered in Flickr®. In other words, language mirrors reality with a fairly high degree of fidelity.

However, the distribution that emerges from the analysis based on visual contexts (Flickr®) shows peculiarities that do not appear in the two models based on word co-occurrences. In particular, the model based on Flickr® suggests that, in our cognitive system, perceptual and conceptual processes are tightly interconnected, and that the perceptual experiences that we live through our bodies (and that are constrained by our biological limits) are reflected in the way we structure semantic information in our memory. Such mechanisms seem to be well reproduced by the distributional model based on Flickr®: physically similar colors (i.e. colors with close wavelengths) tend to be perceived as salient features in similar experiences, and therefore tend to be tagged in similar contexts, and thus tend to be conceptually similar. This mechanism works less effectively with regards to the semantic information encoded in language: colors that are physically similar are not necessarily used in similar linguistic contexts.

4. General discussion and conclusions

The results of these analyses suggest that the semantic information that we retrieve from language correlates to the semantic information that we retrieve from perceptual experience, but that the first source does not cover all the information that we retrieve from the latter source. In fact, the representations that we can model just by using linguistic

information (as in LSA or DM) differ from those that we can model by using perceptual data (as in Flickr®). Still, the two types of representations are positively correlated because there are redundancies between the information encoded in language and in the perceptual experience (e.g. Louwerse and Jeuniaux 2010). As suggested by Barsalou, shallow linguistic processing demands less cognitive effort than deep conceptual processing achieved through the re-enactment of perceptual states. For this reason, during conceptual processing, linguistic strategies peak first (LASS Theory, Barsalou et al. 2008). Depending on the cognitive task that we need to perform, language-derived information might be sufficient to achieve a satisfactory level of understanding. For this reason, a "blind model" based on solely linguistic information can reproduce human processing of linguistic input with a high degree of fidelity. However, this is true for input cued by linguistic means. On the other hand, processing an experience or an image, as well as performing a task that demands the mental simulation of an entity, seems to activate directly the semantic information that we retrieve from perceptual experiences (Glaser 1992; Santos et al. 2008; Barsalou et al. 2008), which might be missing in a model based only on language.

The proposed method of distributional analysis is designed to investigate whether the semantic information encoded in Flickr® tagsets includes perceptual and conceptual features about given concepts that are not fully reflected in the linguistic contexts where the same concepts occur. This case study demonstrates exactly this point: an inherently perceptual domain such as the domain of colors, which sighted people acquire through perception, can be better mapped in a distributional model based on experiential contexts, rather than in a model based on solely linguistic information. The limits of our conceptual knowledge about colors, deriving from the solely linguistic stream of information (leaving aside the information retrieved from perceptual experience) are also explored in a recent study conducted by Connolly et al. (2007) with congenitally blind participants. The authors highlight how the lack of first-hand experience with colors affects implicit judgments about concepts such as fruits and vegetables. Interestingly, however, this does not affect the judgments about concepts such as household items, suggesting that color might be a salient feature only for specific categories.

Given the peculiarities that characterize the semantic representations emerging from Flickr® highlighted in this explorative study, it might be interesting, as a further development, to compare Flickr®-derived representations to those that emerge from databases of semantic features collected in experimental settings (e.g. McRae et al. 2005; Vinson, Vigliocco 2008). These databases contain standardized feature norms elicited directly from human beings who were asked to describe the content of specific concrete and abstract concepts. Evaluating the semantic representations emerging from Flickr® against the semantic

representations emerging from the speaker-generated features gathered in these databases will also contribute to establishing the distributional hypothesis as a psychologically plausible mechanism for making sense of our experiences.

To conclude, the encouraging outcomes achieved in this study support the idea that the social phenomenon of collaborative tagging should be exploited to a larger extent in cognitive science, for it constitutes a powerful resource of semantic data that might shed light onto some aspects of human perceptual and conceptual processing.

Notes

[i] Trademark used with permission from Yahoo!

[ii] With *concept* (type), I refer to a homogeneous category that represents portions of situations that can be gathered under the same lexical label. Concepts (tokens) manifest themselves in individual situations.

[iii] According to a Yahoo! report from August 2011 Flickr™ is a far-reaching community of 51 million registered users, who upload on average a total of 4.5 million photos daily (source: comScore Media Metrix, U.S., August 2011).

[iv] With *situation/episode*, I refer to an individual experience, a single experiential context from which we retrieve semantic information about the involved concepts.

[v] For a review, see Thaler et al. 2011.

[vi] By *semantic features*, I refer to all the range of associations that are stimulated by a given concept: its internal and external properties, the linguistic contexts in which it occurs, the emotions that it arouses and so on. In this perspective, concepts might act as semantic features for the definition of other concepts.

[vii] A *tagset* is a group of tags attributed to a single photograph.

[viii] With *perceptual domain*, I refer to a cluster of concepts that share a good number of perceptual features that make them different from other concepts.

[ix] Freely downloaded from the following link: code.google.com/p/irrational-numbers/downloads/detail?name=FlickrSearch-1.0b.zip.

References

Baroni, M. and Lenci, A. (2010). Distributional Memory: a general framework for corpus-based semantics. *Computational Linguistics* 36 (4): 673-721.

Barsalou, L.W., Santos, A., Simmons, W.K., and Wilson, C.D. (2008). Language and simulation in conceptual processing. In M. De Vega, A. Glenberg and A. Graesser (eds.), *Symbols and embodiment: debates on meaning and cognition.* Oxford: University Press. pp. 245-283.

Barsalou, L.W. (2012). The human conceptual system. In M. Spivey, K. McRae, and M. Joanisse (eds.), *The Cambridge handbook of psycholinguistics*. New York: Cambridge University Press. pp. 239-258.

Bouma, G. (2009). Normalized (Pointwise) Mutual Information in Collocation Extraction. In C. Eckart de Castilho and Stede (eds.), *From Form to Meaning: Processing Texts Automatically, Proceedings of the Biennial GSCL Conference 2009*. pp. 31–40.

Bruni, E., Tran, G.B., and Baroni, M. (2011). Distributional semantics from text and images. *Proceedings of the EMNLP 2011, Geometrical Models for Natural Language Semantics Workshop*. pp. 22-32.

Connolly, A.C., Gleitman, L.R., and Thompson-Schill, L.S. (2007). Effect of Congenital Blindness on the Semantic Representation of Some Everyday Concepts. *Proceedings of the National Academy of Sciences*. pp. 8241–6.

Feng, Y. and Lapata, M. (2010). Visual information in semantic representations. *Proceedings of the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. pp. 91-99.

Firth, J. R. (1957). *Papers in Linguistics 1934-1951*. London: Oxford University Press.

Fodor, J.A. (1975). *The Language Of Thought*. New York: Crowell.

Glenberg, A.M. (1997). What memory is for. *Behavioral and Brain Sciences* 20: 1–55.

Glenberg, A.M., and 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.

Harnad, S. (1990). The Symbol Grounding Problem. *Physica* 42: 335-346.

Harris*, Z. (*1954*).* Distributiona*l* structure. *Word* 10 (23): 146–162.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea.* Cambridge: Mit Press.

Hunt, R.W.G. (2004). *The Reproduction of Colour* (6th ed.). Chichester UK: Wiley.

Izmailov, C. A. and Sokolov, E. N. (1992). A semantic space of color names. *Psychological Science* 3 (2): 105-110.

Landauer, T.K. and Dumais, S.T. (1997). A solution to Plato's problem:

the Latent Semantic Analysis theory of the acquisition, induction and representation of knowledge. *Psychological Review* 104 (2): 211-240.

Louwerse, M.M. and Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition* 114: 96-104.

McRae, K., Cree, G. S., Seidenberg, M. S., *and* McNorgan, C. *(*2005). Semantic feature production norms *for a* large set *of* living *and* nonliving things. *Behavioral Research Methods, Instruments, and Computers* 37*:* 547-559*.*

Miller, G.A. and Charles W.G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes* 6 (1): 1–28.

Pecher, D. and Zwaan R. (eds.) (2005). *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*. Cambridge: University Press.

Pylyshyn, Z.W. (1984). *Computation and cognition: Toward a foundation for cognitive science.* Cambridge, MA: MIT Press.

Sahlgren, M. (2006). *The Word-Space Model*: *Using distributional analysis to represent syntagmatic and paradigmatic relations between words.* Stockholm: University Press.

Steels, L. (2006). Collaborative tagging as distributed cognition. *Pragmatics and Cognition 14 (2): 287-292.*

Thaler, S., Simperl E., Siorpaes K., and Hofer C. (2011). *A survey on games for knowledge acquisition*. STI Technical Report.

Vigliocco, G., Meteyard, L., Andrews, M., and Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition* 1 (2): 215-244.

Vinson*, D.P. and* Vigliocco*, G. (*2008*).* Semantic feature production norms for a large set of objects and events*. Behavioral Research Methods* 40 (1): 183–190.

Wu*,* L.L. and Barsalou*,* L.W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica* 132: 173-189.

Zwaan, R.A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (ed.). *The psychology of language and motivation*. New York: Academic Press.