Linguistic Relativity and Flexibility of Mental Representations: Color Terms in a Frame Based Analysis

Leda Berio

Abstract This paper connects the issue of the influence of language on conceptual representations, known as Linguistic Relativity, with some issues pertaining to concepts' structure and retrieval. In what follows, I present a model of the relation between linguistic information and perceptual information in concepts using frames as a format of mental representation, and argue that this model not only accommodates the empirical evidence presented by the linguistic relativity debate, but also sheds some light on unanswered questions regarding conceptual representations' structure. A fundamental assumption is that mental representations can be conceptualised as complex functional structures whose components can be dynamically and flexibly recruited depending on the tasks at hand; the components include linguistic and non-linguistic elements. This kind of model allows for the representation of the interaction between linguistic and perceptual information and accounts for the variable influence that color labels have on non-linguistic tasks. The paper provides some example of strategy shifting and flexible recruitment of linguistic information available in the literature and explains them using frames.

Keywords Colors · Labels · Concepts · Perceptual information · Frames

1 Introduction

Cross linguistic[1] research about basic color terms has been for a long time a central concern in the debate regarding Linguistic Relativity, i.e. the influence of language on conceptual representations. However, this has been seldomly connected to the

L. Berio (✉)
Department of Philosophy, Heinrich–Heine–Universität, 40204 Düsseldorf, Germany
e-mail: leda.berio@hhu.de

issue of the *structure* of mental representations. In this paper, I will argue that a frame-based model of mental representations allows for the representation of the relation between the perceptual information contained in color concepts and their linguistic labels in a way that is compatible with the empirical evidence used in the Linguistic Relativity debate. In doing so, I shift the problem of Linguistic Relativity to a matter of the structure of mental representations. In the account I present, mental representations are conceived as complex functional structures that are dynamically and flexibily recruited according to the task at hand and that include both linguistic and non-linguistic information. The core claim of the paper will then be that such a model allows for the presentation of the interaction between different components of a mental representation and can account for the variable influence of linguistic labels on color-related tasks in terms of strategy shifting and flexible use of mental representations' components.

In the first part of the paper, I delineate the debate about Whorfianism and its more recent declinations, connecting the debate to the problem of flexibility in mental representations. Secondly, I briefly present a few examples of effects of what is called "shallow Whorfianism", describing the available experimental evidence. In the third section, I propose a way to represent color concepts in frames and I subsequently show how this can be applied to concepts in general. In Sect. 4 of the paper, I explain how this view can be fruitfully applied to communicative situations and pragmatic effects and, most importantly, to model the experimental data presented in Sect. 2. In Sect. 5, I provide an example from a different conceptual domain (number representation) that can be treated efficiently with the proposed model. In Sect. 6, I show how, in the same spirit, the model can be used to model a classical color task, i.e. the Stroop task. Finally, I draw conclusions regarding the debate and suggest further necessary steps.

2 Color Terms and Whorfianism: Some Coordinates

2.1 Universalism, "deep" and "shallow" Whorfianism; Intertwined Issues

For a long time, the debate regarding color terms acquisition has been influenced by a (sometimes well grounded) bias against the idea of Linguistic Relativity: one of its earliest formulations, namely the Sapir-Whorf hypothesis, suggests as a matter of fact a particularly strong and simplistic influence of language on thought. However, the debate has seen a partial re-ignition due to more modern studies and techniques that, revisiting the Whorfian hypothesis' too strong initial assumptions and statements, have postulated a role for language in various tasks. This is also partially due to the fact that what was initially taken as the final word on the color terms debate (namely the study by Berlin and Kay 1969) has been scaled down to be an important but not decisive piece of evidence. This is not the place to discuss Berlin and Kay's

research and proposal for universal patterns in color terms; for the present purpose, it is sufficient to keep in mind that it is possible to postulate some kind of influence of color terms on color cognition without necessarily contradicting Berlin and Kay's fundamental insight that there are universal tendencies and/or constraints on focal colors that are perceptually more salient and therefore easier to identify in absence of corresponding color terms.

It is essential to specify that this debate is concerned with a particular aspect of language, which is indeed lexical labeling: most studies regarding color cognition are focused on whether or not color terms that are present in one language have any influence on performance as far as color recognition is concerned. This brings us to the other important specification, which is that the debate is concerned with influence on perception and categorization tasks. The color words debate is often enough considered the privileged (if not exclusive) ground for deciding about the whole debate concerning Whorfianism and Linguistic Relativity. However, it is worth underlining that the main focus of a big part of the debate is very specific: whether or not lexical entries influence perception and attention mechanisms.

As a matter of fact, as Lalumera (2014) already notices and as it will be clear in the next paragraphs, the evidence available in the literature cross-cuts the distinction between Whorfianism and Universalism, since there are in this sense various kinds of results suggesting, on the one hand, some influence of linguistic labels on perception mechanisms, and on the other hand, rejecting the extreme claim made by language relativity supporters in the past, namely that language strongly shapes mental representations. Thus, the distinction between Universalism and Language Relativism has partially been replaced in the literature by what Lalumera phrases as a distinction between "deep" and "shallow" Whorfianism, separating those phenomena where the influence of linguistic labels seems to be constant, pervasive and stable, from those cases in which it is "only" a flexible, context dependent, task dependent influence of some sort. The reason why this distinction cross-cuts the previous one, i.e. Universalism vs. Whorfianism, is that the old debate was concerned with a less fine-grained question: through the universalist lenses, Whorfianism was seen as threatening the idea of concepts as something that follows potentially the same "rules" of formation and development regardless of the language of the speaker, therefore menacing the idea that humans have a somehow universal conceptual repertoire. Whorfianism, on the other hand, was concerned with the fact that universalism seemed not to admit any interference of language with mental representations' structure and complexity. Framing the debate as "deep" and "shallow" Whorfianism shifts the focus of the debate to a somehow more pragmatic issue, namely how do linguistic processing and linguistic labeling interfere with non- linguistic processes, including but not confined to conceptual formation, and to what extent is that relevant in non linguistic tasks. The question then becomes, when is this influence relevant and how stable and pervasive is it. In what follows, I will also try to argue that this might shed some light on how to think of conceptual structure itself, without making the bold, original Whorfian claim that language invariably shapes representations.

Note that this whole debate is better understood if connected with the parallel but distinct issue regarding cognitive penetrability.[2] Cognitive penetrability can be defined as the property of perceptual experience to be influenced by what happens at the so-called higher cognitive level; in other words, we speak of cognitive penetration when perceptual experience is influenced by beliefs, desires, intentions and concepts (Newen and Vetter 2017). In a way, the debate can be conceived to proceed hand in hand with the issue treated here: admitting an influence of linguistic information on non linguistic processing means admitting permeability of perceptual experience. The problem of permeability, on the other hand, is of a broader nature, as it comprises considerations regarding modularity and specialization of brain areas; in other terms, the debate regarding permeability brings us to a broader scale of issues regarding cognition in general. The focus of the current paper is on the relation between linguistic labels and color *concepts*; which means, on the one hand, that perception is obviously relevant for the discussion, given color perception is at the center of the debate; but also, on the other hand, that the focus is already on mental representations employed in experience and not on perceptual experience itself, which implies that the focus is on the level of "higher cognition" only.

Admitting permeability means admitting that the *experience* of color changes depending on (among other things) linguistic processes; the debate regarding Linguistic Relativity focuses on whether or not the *concepts* related to color and used in perception are influenced by color labels. This claim is therefore both weaker and related. Related, because color mental representations are supposedly recalled in color perception; but weaker, because it moves prevalently at the level of higher cognition (linguistic information influencing representations) and because it does not make claims on the *experience* related to color but only on the representational means employed.[3]

As it will be clear in the rest of the paper, the view proposed here, despite being mainly concerned with mental representations and higher cognition as said, assumes permeability. As a matter of fact, it is assumed here that different kinds of information such as perceptual and motor information are integrated in mental representations along with more abstract kinds of information, like linguistic-based one. In this sense, the view even *endorses* an account of mental representations that accepts cognitive penetration and refuses strict modularity.

Getting back on the shallow–deep spectrum, "deep Whorfianism" is problematic to argue for, given the scarce evidence in favour of an influence of language on thought that actually is not task dependent but stable and pervasive. Moreover, it is arguably a type of influence that is more likely to be related to words and concepts that are more complex and less perceptually-bound than color ones, as it will be argued elsewhere.[4] However, the focus of this paper is the so-called "shallow" Whorfianism, or, in other

---

[2]Thanks to the anonymous reviewer for pointing out the necessity of mentioning this.

[3]Note that Macpherson (2012) contains an interesting review of color literature connected to cognitive penetration.

[4]One assumption of my work on the interface between language and cognition is that it varies depending on the type of concept/category that is considered.

words, the influence of language that is only detectable in specific tasks. In the frame of the Universalism-Whorfianism debate, this kind of influence is irrelevant, because the question at issue is whether having a different language irreversibly shapes the conceptual repertoire in a deep, pervasive way. In this sense, the answer going along with shallow Whorfianism is, clearly, negative. However as Lalumera points out.

> […] some Whorfian effects show themselves to be task dependent and temporary. A question on this point is worth raising here. Is that enough to deem such effects as uninteresting, qua task dependent and temporary? The answer is that it would be enough, but at the price of committing to the view that only stable and context-free representations are employed in perception and cognition. (p. 7).

This is an essential remark: arguing against any kind of influence of language on non-linguistic cognitive processes appealing to the fact that the supposed influence might only be task dependent and not *always* present means endorsing a view of mental representations that is not trivial (anymore). In other words, it means committing not only to the idea that there is a stability in mental representations and categories, but also that this stability is such that everything that regards the flexible, online, task dependent application of these same categories is not relevant because it does not tell us anything about mental processes. Lalumera points out that this does not seem to be the case, and that there is plenty of evidence suggesting the contrary. My claim goes in a slightly different direction: I think that what the evidence available in the literature suggests is that a way to represent the interaction between linguistic labels and conceptual units is needed and that, whatever the model, it has to cope with how variable this influence actually is. In what follows, I will briefly present some examples of "shallow Whorfianism" that are present in the literature and then propose a way to model them using frames. I will then try to show how the model can be flexible and fruitful in dealing with some challenges that conceptual representations and language present to us, if we assume a view of representations as flexible adaptable structures that can be differentially activated depending on the task at hand.

2.2 "Shallow" Effects of Color labelling

Many examples in language cognition and color deal with perception tasks. In this paragraph, I will focus on two well-known studies that are often referred to in the literature because they're considered evidence that Whorfian influence is "shallow" because it is task dependent. Later in this paper, I will focus on one of them as a paradigmatic case that points in the direction of a flexible, context dependent use of linguistic representations in non-linguistic tasks, while at the same time underlining the open questions that are left.

A well known and cited study, therefore worth mentioning as a valid example, is Winawer et al. (2006). Russian has an obligatory distinction between *light blue* and *dark blue* (*goluboy* and *siniy*), as many other languages, like Greek and Italian, do.

In the study, subjects (divided between Russian speakers and English speakers) were shown three color squares arranged in a triad; the task consisted of saying which one of the bottom squares was identical to the one on top, while reaction times were measured. In "within category" trials, the square was from the same color category of the match, whereas in "cross-category" trials the distracter and the match belonged to different categories in Russian color categorization system.

The hypothesis was that the presence of a color boundary available in one language (Russian) but not the other (English) would have affected performance across the boundary; more specifically, that Russian speakers would have made faster cross-category discriminations than within category ones. The prediction was confirmed: there was indeed a difference between the performance of Russian speakers and that of English speakers. Even more interestingly, the effect disappeared if the subjects also had to perform a verbal interference task at the same time (the task consisted in silently rehearsing digit strings): it seemed, then, that blocking language resources with task-irrelevant processing was preventing the effect. At the same time, estimating the difficulty of the trials, the research group found out that the difference between cross-category and within-category trials performance for Russian speakers increased the more difficult the discrimination was.

Several interpretations can be given of the results. First of all, the fact that the facilitation disappears when linguistic interference is added, suggests at least two things: firstly, that the effect on perception is temporary and tied to the specificity of the task, and secondly, that language labels are extremely likely to be the cause of the effect, because linguistic coding seems to be involved. Clearly, then, we are in the realm of what has been referred to as "language as a meddler" (Wolff and Holmes 2010): there is an *online* interference that takes place during a certain task and that is heavily dependent on the context and conditions of the task itself. It is also clearly a case of language *changing* the performance as far as an already existing skill is concerned, namely, to be precise, color discrimination. One of the most interesting results is definitely that the difference in performance increased if the task was perceptually more difficult: this suggests that language was used as a facilitator of some kind, with linguistic labels possibly used too, as a support for the difficult discrimination task. In this case, then, we have a case in which language is *improving* the performance on a task.

Different kind of data comes from studies like that of Roberson et al. (2008), who explored differences between English and Korean speakers. Korean has fifteen basic color terms, as opposed to the eleven English ones. Once again, color perception was the focus of the study, which was aimed at comparing linguistic distinguishability and perceptual one. It is often argued that language centres are to be located on the left hemisphere and categorization functions are to be attributed to clusters in the right hemisphere; wanting to test this distinction, the study investigated the categories of *yeoundu* and *chorok*, respectively *yellow-green* and *green* in Korean. In the task, participants were presented with an array of color patches, among which one was different from the others. The patches all belonged to the category *green* for English speakers; for Korean speakers, however, the "odd ball" patch could belong either to the same category as the others or not. Participants had to say whether the odd ball

was right or left in the screen (hence, the stimulus was presented to be elaborated either in the right or in the left hemisphere). Once again, there was a difference in cross-category and within-category discrimination: Korean speakers made faster cross-category judgments compared to within category ones; the effect was present regardless of the visual field. However, a comparison between fast responders and slow responders led to an interesting result; fast responders only were facilitated when the stimulus was presented in the right visual field, whereas the effect was present for slow responders even for the left visual field-presented stimuli. This was interpreted as a sign that the effect was due to linguistic labels: in case of slower responses, time allowed the information to be transmitted via *corpus callosum*. Even here, the influence of language labels is evident, but at the same time clearly dependent on task constraints. Similarly to the previous case, moreover, we are talking about an influence of language labels on perception and attention mechanisms.

In both the mentioned cases, there is an influence of language that is clearly constrained by determined conditions and tasks: moreover, these are not isolated cases. Evidence very similar to Roberson et al., for instance, was collected by Gilbert and colleagues (2007). In general, what this kind of evidence tends to suggest is that influence of color words is variable and task dependent, and this seems to be suggested by other studies as well in other semantic domains (see Papafragou, 2008 for instance). However, these results, while suggesting cognitive penetration of some kind, still do not shed any light on what the possible relation between linguistic labels and mental representations is and how it can be modeled.

3   Frames and Representation of Colors

Let us take a step back and consider the kind of picture that is compatible with the presented data. As underlined, this kind of data is often cited in the domain of Linguistic Relativity as an influence of language on color concepts; however, little is said about how color concepts enter the picture.

There are several accounts out there that try to tackle the issue of the structure of mental representations, and this paper is not meant to be a review of them; on the other hand, it is at least worth underlining that papers as influential as the one published by Casasanto and Lupyan (2019) efficiently sum up plenty of good evidence in favour of representations as task and context dependent in various ways, showing how evidence from psycholinguistic and cognitive science accounts for a great flexibility in mental representations.[5] In what follows, I will adopt the idea that concepts can be efficiently represented as frames as developed by Barsalou (1992). There exist several theoretical elaborations of frame theory and the research regarding its compatibility with other theories of mental representation is vast; for the purpose of the paper,

---

[5]Casasanto and Lupyan use this evidence to argue, at the same time, against (1) the idea that there is any stability in mental representations (2) the possibility of talking about *shared* representations. I think their claim is, in this sense, far-fetched, but this goes outside the scope of this paper.
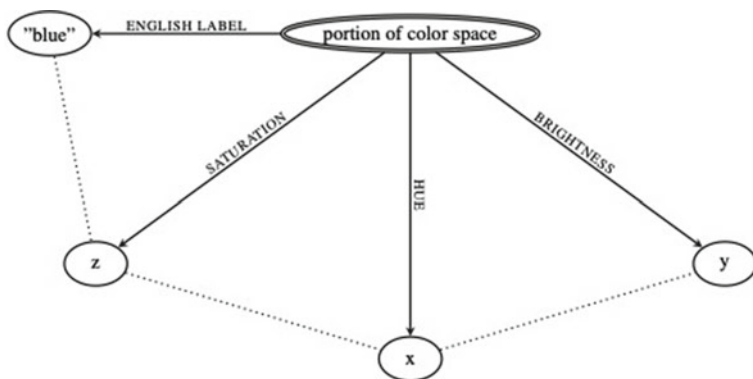
Fig. 1 Frame for the color concept BLUE

however, only a few specifications are needed, starting from the idea that frame theories assume that an efficient way to describe and model conceptual components is to think of complex structures where attributes get assigned unique values.

Furthermore, note that frame theories are quite different from feature lists approaches, for instance, or from concept atomism, since they all assume that concepts have a fine-grained complex structure (contra atomism) and that attributes are functional, contra feature list approaches.[6] However, choosing frames as a model, in this instance, does not mean necessarily buying one specific philosophical theory of concepts. Assuming this is a good model for conceptual representations does not mean necessarily take a stance on the issue, for instance, of whether or not prototype theory is a good account for concepts; there is currently a lot of research regarding how and when frame theory can be integrated in other approaches, and that heavily depends on the kind of frame theory that is chosen. For the purpose of this paper, however, only two characteristics of frame theory have to be assumed: the possibility of building recursive structures (1) and the possibility of imposing *functional* relations and constraints among attributes and nodes (2).

Let us assume that labels for colors can be considered as an attribute, *label*, functionally connected to another node in an attribute-value structure.[7]

The frame for a color concept then would look like Fig. 1. The expression "portion of color space" is here intended as a place holder for a region of the color space, i.e. a value interval (note that thinking about it in terms of a prototypical blue or an exemplar-like blue does not make a difference for the present purpose). The arrows in the frame represent the functional attributes; the non-arrow arches represent constraints between the attributes. Roughly speaking, the idea is that a color concept can be represented in terms of a portion of color space characterized by a given

---

[6]This is a characteristic of Düsseldorf frame theory, adopted in this paper; see Löbner 2015.

[7]Modelling the relation between linguistic information and conceptual one, far from being contradicting frame theory, is also the focus of other current research. For a compatible account see for instance Beckmann, Petersen and Indefrey, submitted).

saturation, *hue* and *brightness*, whose value range constraints the attribute *English label*. Ideally the constraint can be spelled out in these terms:

$$If (x \in \{..\}, y \in \{..\}, z \in \{..\},) then \iota = \text{``blue''} \tag{1}$$

where ι represents the value of the attribute *English label*, which is in this case "blue". The formula reads so that, if the values of *hue*, *brightness* and SATURATION are included in a given interval, then a given label applies to the portion of color space considered.

Note that there is a clear difference between attributes like *hue*, *brightness* and *saturation* and one like *English label*. In the first case, we have information whose knowledge does not have to be declarative, whereas in the latter we have a linguistic attribute of which we necessarily have a declarative knowledge. This is not problematic because the frame does not represent the declarative knowledge about a color, but rather the structure of the representation. This applies even more significantly to the values that the attributes take, since it might be explicit in my representation that colors are characterized by these three aspects, but I might not know the values involved. Clearly, the idea for these three attributes is that the values they take range in a determined interval. The importance of specifying the language considered should be clear; the idea is that different languages will have different constraints operating (constraints where the intervals for the values of hue, brightness and saturation are different) and will give different results in terms of the label. Another obvious necessity of specifying the language in the attribute will be, for instance, considering the fact that bilingual speakers might have more than one label available for the same values x, y and z. Such a mental representation, then, contains both explicitly known and implicitly known information, represented by values that can be either an interval or not, depending on the kind of attribute.

Let us embed a frame for a color concept like this one in a different frame, in Fig. 2. The given example illustrates a frame for the mental representation of a banana. Clearly much more than what is represented could enter a speaker's representation of a banana, but only salient or situationally-relevant attributes are listed in the representation. The underlying idea is that this might be a way to represent what an individual speaker has in mind when thinking about a banana.[8] Clearly, an assumption here is that the linguistic label for an object, like for instance a banana, is part of the set of information connected to the perception of the object in the mind of the speaker or, in other words, that it makes sense to think about the semantics of word meaning not to be disconnected from mental representations of the objects that words denote. The advantage of such a move will hopefully be clear once we will be proceeding with the rest of the argument.

---

[8]Albeit, again, with all the simplifications applied here for the sake of brevity. The individual's representation of a Banana might include a lot of idiosyncratic information: judgements about how bananas taste like, for instance, or individual experiences concerning this type of fruit, or even some kind of danger signal in case of an allergy to bananas. The amount of idiosyncratic information included in a frame is a matter of discussion.
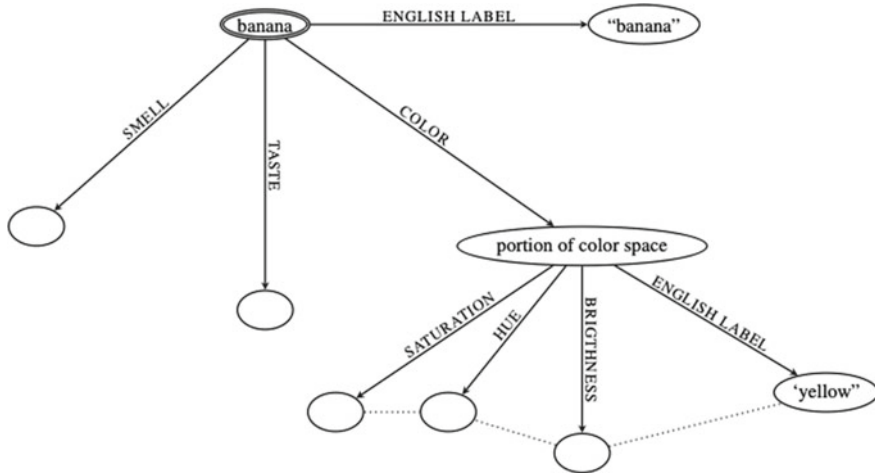
Fig. 2 Instantiated frame for a banana

First thing to notice is that the frame includes information that is basically only perceptual in one of the nodes.

The idea is that a flexible structure like a frame (or, better, the interaction between frames) can be used to incorporate different sources and kinds of information, including purely perceptual one. The intuition under this frame is that different essential features of "banana" are listed that constitute some of the relevant parts included in an individual's representation of what a banana is. Other standard attributes we probably might associate with it include, for instance, SHAPE. COLOR is also a standard attribute; what is fundamental here is that frames are recursive, combinable structures. In this case, the color of a particular banana the speaker might have in mind is related to the concept of that color, which might be an exemplar-like representation or a prototype, for example. This concept is then labeled in English. Just like in the "banana" case, the label is considered an attribute among others in the mental representation. The suggestion, then, is to consider the fact that an attribute like *English label* can be inserted and that it applies to both the color and other features of the frame.

Note, furthermore, that the frame represents the banana in the context of ripeness; it is clear that in another context the value for the functional attribute COLOR could be a different portion of the color space (since, for instance, we would have a brownish color when seeing a overripe banana, or a greenish color when seeing one that it's not ripe enough). In that case, the values for the attributes *saturation*, *brightness* and *hue* will be different, and depending on the constraints operating on the language, the resulting label will be different.

Now, one of the advantages of frames is that they spell out the functional relationships between elements of the representations and, therefore, can be used to give a

picture of what happens during communication in an effective way. In the next session, I will briefly discuss two kinds of communicative phenomena that can involve color words.

4　Color Words and Flexible Use of Representations' Features

A characteristic of communication involving color words is that it can give rises to interesting phenomena; to proceed with the argument, let us consider some of the most common examples that can be given when treating the sorites paradox or models of vagueness (see for this variant Rayo 2011). Having a grayish-blueish house among a group of houses that are painted in red and green, we can successfully utter.

[1] Peter's house is the blue one.

and be understood as indicating the grayish-blueish house. In this context, the portion of color space the color of the house can be placed in can be labeled correctly.

However, in a context where the block consists of a blue house, the same blueish-grayish house, a red house and a green house, [1] cannot be used to point to the second one. In this case, "blue" does not apply correctly (or, at least, it does not represent the most successful communicative choice), even if we are considering the same portion of perceptual space. In other words, the label we are using in communication has to change to make the conversational exchange effective. The value of the attribute, then, will vary.

Integrating the two frames representing the two houses can help (Fig. 3); the strategy of labeling the grayish house (house number 2, for instance), "blue" is not a felicitous one because it means recalling the same label used for house number 1; given that the task includes differentiating between the two houses, having the same label does not aid the discrimination and it's therefore not a winning strategy, communicatively speaking. In this context, the discrimination task cannot succeed because the label can be applied to both houses. The frame representation makes the pragmatic effects, in this way, very easy to spot.

The first type of variability I want to draw attention to is therefore this one; color labels for the same portion of color space referring to the color property of an
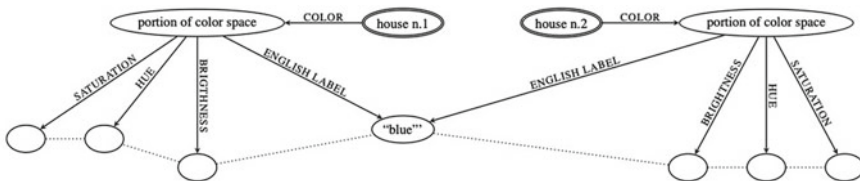

Fig. 3 Two houses' frames

object vary in their communicative efficacy. It is essential to stress that this is a point regarding how mental representations are *used* in *communication*. It is certainly true that, giving an array of color terms available and wanting to apply them in a rigorous way to a representation of color space, we do not have the same kind of phenomenon, but rather a series of determinable-determinate relations: hence, a portion of color space "blue" that can be labeled, on a more fine grained level, "ultramarine" and another that can be "Nivea blue".[9] However, what is meant with the given example is something different, i.e. that a communicative situation can make a label for a determined color more or less communicatively efficient and appropriate in a context, even more so in Sorites-like cases, where this depends on whether or not the perceived color is close in perception to other present portions of the color space. Frames make it particularly easy to see, granting a format of mental representation modeling that aids the understanding of pragmatic effects.

There is also another element of variability, namely the relevance that the activation of a determinate attribute (and therefore of the respective value) has in a determined situation. In other words, at least as far as a certain understanding of frame theory is involved; attributes can be activated or not during tasks that involve the representation in question. Let me use another example at the intuitive level to express the idea. Let us assume I ask a colleague to hand me a folder in my office that contains the notes from the Dynamic Semantics class I am following. The colleague knows me and my office and knows that my folders are all of the same color, say gray, and therefore to find the right folder she will have to read the tags until she finds the one that says "Dynamic Semantics" and then give me the folder. In this case, information about color is not relevant for the task that my colleague has. Let us now imagine that, in the exact same dialogical situation, my folders are colorful, and that my colleague knows my "Dynamic Semantics" folder is the red one; browsing through my shelves in my office, she'll look for the red folder; color information will be in this case salient for the task at hand. This has a lot to do with the fact that the color of an object can be of some relevance or not depending on the situation. When browsing the room looking for an object, different characteristics can be relevant and therefore acquire salience.

There's no intention here to directly compare a perceptual task like that described in the study of Winawer and colleagues to the described situation; the two tasks clearly involve different levels of explicitness and entail different relationships between the attribute color involved and the rest of the representation; however, the point is to embrace the intuitive idea that information about certain features of a determined object can be more or less salient and relevant depending on the task at hand. What these classical examples in pragmatics show is that, in communication, features associated with an object can acquire relevance and salience depending on the situation at hand. In these communicative situations, arguably, mental representations are employed to "solve" the comprehension or production task. In the case of the red folder, different attributes acquire relevance.

---

[9]Thanks to the anonymous reviewer for bringing my attention to this fact.

This kind of idea is not only intuitively plausible, but also what underlies research enterprises in psycholinguistics that are meant to assess what the relationship between concepts and their components is; for instance, studies like Redmann and colleagues (2014) investigate the activation of color attributes in high color-diagnostic concepts (like, for instance, bananas). Studies like this focus on language production; however, the idea is that concepts can be treated as complex structures whose different components can be "activated" depending on the situation. Moreover, it is assumed that definite relations among attributes and nodes in a frame exist, the idea being that the activation of a conceptual component can potentially facilitate the activation of other parts of the concept.

Another analogy will help clarify the position. Consider my own representation of DOG. Presumably, it entails different kinds of attributes encoding several kinds of information - purely perceptual, verbal, and so on. Approximately, a frame representation of DOG for me might include not only information about basic dog attributes - such as for instance number of legs, fur, eating habits, and so on, but also plenty of information about Nala, my dog, about other dog encounters that I had in the past, about my grandma's dog that I got to know when I was very young, about the names for dogs I've heard most often when in Italy, and so on. This entire repertoire of information, however, does not need to be recruited every time I have to activate my dog representation in a communicative situation; it's reasonable to think, on the contrary, that this only happens when certain kind of information is required, or relevant, for a given task - namely, the one I am performing, whatever this might be. Depending for instance on the communicative situation, I will need to recruit different kinds of knowledge.

Let us now apply this understanding of concepts and attributes within them to the main focus of the paper, trying to put the pieces together. The debate is open as far as how lexical information enters the conceptual domain, as described above; the question of how linguistic representations and non-linguistic ones interact is precisely the kind of question that, after all, guides the debate about Linguistic Relativity. On the other hand, if one assumes that information about how certain perceptual features can be linguistically coded in different ways (hence, that we can assume the presence of attributes-like structures like the LABEL one and that the value can change) and that conceptual components can be recruited *according to the situation and the context at hand*, it is natural to assume that the linguistic information can or cannot be activated and recruited, depending on the context. The modalities and circumstances of this activation, then, would need to be investigated.

A case like that of Winawer seems to suggest that conceptual representations of colors, and consequently their labels, can be used and activated during a perceptual task; one of the possible interpretations of the results is that, while English speakers operate comparing different perceptual inputs without activating linguistically coded representations, Russian speakers use a different strategy, namely they employ color concepts and their labels; at least that's what seems to be suggested by the difference in performance. Crucially, however, this kind of strategy seems to be replaced by the same strategy English speakers employ, in case of linguistic interference: somehow, then, performing another linguistic task "blocks" or inhibits the label-influenced
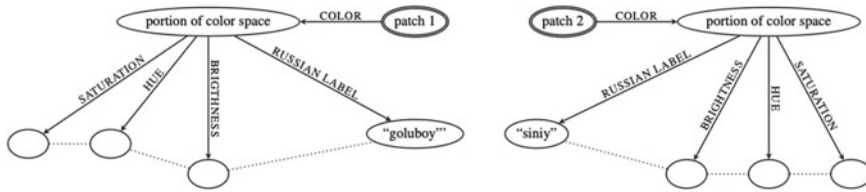
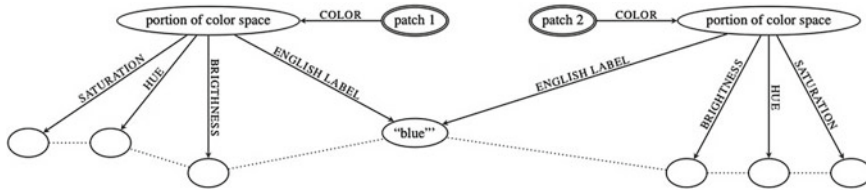**Fig. 4** Winawer's task in frames: Russian

Fig. 5 Winawer's task in frames: English

strategy. Given the fact that the task is still possible for English speakers, this is clearly not something that prevents them from performing the task, regardless of the presence of color labels. What this study seems to suggest, then, it is that recruiting or not recruiting linguistic information can depend on the type of task: in this sense, the choice of strategy is flexible.

Let us try and represent this in frames again with Figs. 4 and 5.

A plausible explanation that is easily representable in frames is that the task is solved by the Russian speakers by comparing two different nodes including linguistic information. This strategy is not available in the case of English speakers, since there is only one node containing linguistic information available; therefore, a strategy based on comparing, for instance, visual patters in SATURATION, HUE and BRIGHTNESS is used. Russian speakers can then shift to the same strategy when the label attribute is unavailable- i.e. in within-category trials.

To reiterate: this means assuming that it is possible to draw a parallelism between concepts like BANANA and concepts like BLUE; in other words, assuming that it makes sense to consider an attribute like *label* (in language x) to be something that pertains to the representation of both. In a sense, this is the first tenet of the model presented here. The second tenet is that a mental representation can be considered as a structured file where not every part gets activated every time the concept is evoked; instead, the amount and the kind of information that will be used in the task at hand will vary according to task constraints, context and possibly other factors. Finally, a point that has been stressed while presenting the view is that different kinds of information, of perceptual and not perceptual nature, can be incorporated in the same mental representation.[10]

---

[10]This is clearly not the only available theory. An alternative account can for instance be found in Newen (2011) A thorough comparison between the two views would be fruitful but would go beyond

Arguably, more research has to be done in this direction, as the issues are multiple and complex. However, it should be clear that results of studies like that of Winawer or Roberson should be considered as interesting because they fit into an account of cognitive processes manipulating representations in a flexible, task dependent way, where different information is recruited according to what is useful for the task at hand. In Winawer's case, paradigmatically, linguistic labels seem to play the role of facilitators for the task at hand, or at least to make a difference when recruited. Phrased using the vocabulary introduced until now, this implies assuming that there are complex interactions among linguistic information and perceptual information which are functionally connected and can be differently employed. Frames are just one way to represent this kind of relation: however, they help in seeing how data such as that presented, more than settling the debate about language relativism, should suggest to see it in another light. A difference between "shallow" and "deep" Whorfianism ceases to be relevant, once one assumes that the kind of information that has to be considered when modeling mental representation can be of different kinds (linguistic and perceptual, for instance) and that this kind of information interacts in complex ways: the fact that effects of language categorization on cognitive tasks vary depending on context and task demands seems to point towards an understanding of mental representations precisely in this direction.

So far, it has been argued that a view of mental representations that involves flexible use depending on the task at hand can be represented efficiently in frames and that it has a good chance to be related to a model of how representations are used in communication. However, a few steps are still needed. In the Russian-English speakers example, what we apparently have is the use of two different strategies for performing the task: however, there is still no direct evidence in favor of considering "LABEL" as an attribute that gets activated depending on the task. For all we know, the strategy employed by English speakers (and by Russian speakers when linguistic interference is present) might not include any kind of conceptual activation. Participants might be comparing perceptual input, solving the task on the basis of this comparison, and using a strategy based on labeled mental representations instead when two different color terms are present: this suggests switching between strategies, but does not support necessarily the idea that the linguistic information in a concept can be activated or not depending on the situation. I think this is a viable option, as will be argued below. In order to push further Lalumera's suggestion, to consider the compatibility of the color terms evidence with a more dynamic picture

---

the aim of the paper. Two basic differences are however to be noticed; firstly, Newen adopts a model where relations between conceptual parts are not spelled out in terms of functional relations like in frames. Secondly, he makes a distinction between two different concepts: RED referring to the property of being red and RED EXPERIENCE referred to the property of having a red experience, where the information contained in the first can be integrated in the latter, albeit not as a defining component. I believe this idea could be integrated in a frame network, but this would require further investigation.

of mental representations, it is necessary to go a few steps further. To get there, we will consider now a different example from another conceptual domain before turning to colors again.

5   A Brief Excursus into Another Conceptual Domain: Counting and Motor Representations

As argued so far, in the case of cross-linguistic evidence for color terms, the debate has focused a lot on whether effects are to be considered "just" shallow and temporary or "deeper". In the context of embodied cognition, something very similar has happened, in a somehow opposite direction. Embodied semantics is concerned with the role of motor and perceptual representations in conceptual units, the idea being that is worth exploring the multimodality of mental representations or, in other words, the role that sensory modalities play in their structure, use and retrieval. One of the battle grounds in the embodied cognition debate has always been that of abstract concepts: even if it's more or less accepted that motor and perceptual information can have some relevance as long as concrete concepts are concerned, the same does not hold for concepts that, intuitively, have less to share with perception, hence abstract concepts. Moreover, one common argument against embodied cognition lies in the idea that, even when perceptual and motor resources are recruited during semantic processing, this is only a somehow shallow "cascade effect" that has nothing to do with "deeper" conceptual processing (Mahon and Caramazza 2008).

In the context of research regarding representations of numbers, which are considered quite abstract, there have been several attempts to connect numbers and counting to the more (supposedly) concrete domain of space, the idea being that abstract concepts like mathematical ones are mapped to more concrete representations like spatial ones, which is what guarantees their being "grounded" in experience. In a famous study run by Dehaene and colleagues (2019), the so called SNARC (Spatial Numerical Association of Response Codes) effect was described: large numbers elicited rightward response and small numbers leftward ones, meaning that small numbers were classified faster with the left hand and bigger digits were classified faster with the right hand. Since similar effects were found as long as the vertical axis is concerned (up for bigger digits and down for smaller ones), this kind of idea was investigated in a number of other studies. A particularly interesting one is that by Pecher and Boot (2019). The task was to judge the magnitude of numbers in comparison with other digits: the stimulus was a digit that was located congruently or incongruently with the image schematic location of the number (left for smaller digits, right for bigger ones). In the concrete contexts, participants had to say whether the digit was bigger or smaller than the one in concrete sentences ("The man read two books a day"). In the abstract context condition, the digits were to be compared to other numbers. The idea was to test whether the congruent spatial condition was facilitating the task or not, which ended up being true *only for the concrete context*.
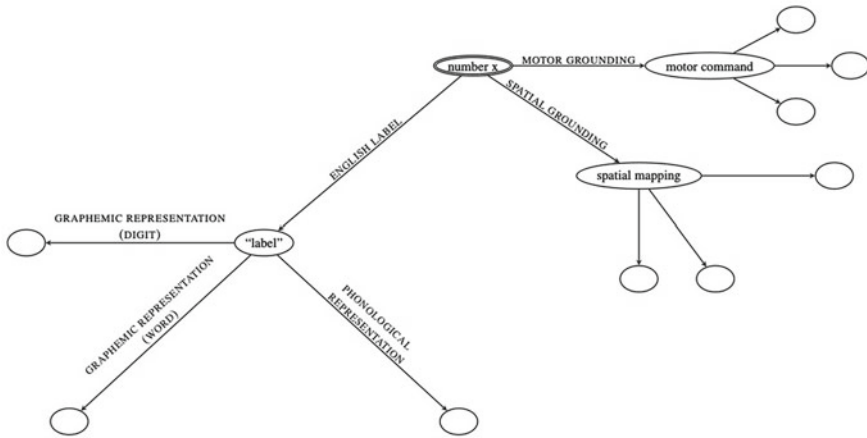
Fig. 6 Frame for a number

Regardless of the debate about embodied cognition, which is vast and complex, the result is interesting because it has been used to argue against the idea that spatial representations are relevant for number processing *because they only appear to be used in certain processing contexts*. This is somehow very similar to what happens in the color labeling debate: even here, the key of the arguments lies in the fact that certain kind of information is only thought to be relevant in determined contexts and tasks. However, this is hardly enough to say that the positive result (the facilitation effect in the concrete condition) is not interesting: on the contrary, it suggests that different processes are going on linking different kinds of information depending on the task at hand. Moreover, the result goes hand in hand with theories of embodied cognition like that proposed by Barsalou (2008), where the role of motor and perceptual representations and that of linguistic ones *varies depending on the type of task*, but where both have a crucial role in conceptual representations.

Let us look at a possible frame for a concept of a number in Fig. 6.

Different kinds of attributes are present, comprising different kinds of information. A number has a label, which implies a phonological representation *and* a graphemic one and, in this picture, includes spatial mapping information and possibly motor grounding (lots of the research regarding grounding of number has focused on finger counting).

A frame like that in Fig. 6 *does not* imply that motor grounding and spatial information are always recruited when the concept of a number is evoked. On the contrary, it is conveniently compatible with the view of mental representations that has been presented so far and with the idea that different attributes can be recruited depending on the situation at hand. Let's consider the experiment reported: in one condition (the concrete one), spatial information seems to be relevant, since the subjects' performance changed depending on whether the spatial information was congruent with the magnitude of the numbers or not. One can then assume that the attribute named

here "spatial grounding" was then evoked and recruited. The same does clearly not apply to the abstract condition: in this case, the spatial information did not seem to be relevant, since the performance did not change depending on the congruency of the position. This, more than speaking for an alleged scarce relevance of the spatial mapping, seems to suggest that some other kind of information was relevant for the task: for instance, the graphemic representation was probably employed. Lacking a concrete context for the digits, the task was performed using a different strategy, which probably included in this case comparing the graphemic representations of the numbers: this is another kind of information, namely visual. Even in this case, there is a switching of strategies. However, this time, it is plausible to think that different parts of the involved mental representations are recruited. Depending on task demands and conditions, different parts of the representations are relevant, and different attributes are activated. The frame captures the multi-modal nature of the concept and the flexibility that underlies its use.

6 Back on Colors: Stroop Task And Language-Perception Interface

Let us then come back to colors now, and consider another set of evidence that is often discussed, namely the Stroop effect. The phenomena was investigated for the first time in 1935 (Stroop 1935), and very often recreated. In the traditional set up, color words are printed in either congruent or incongruent ink (e.g. the word blue is printed either in blue or red, for instance), and participants are instructed to name the color of the ink used for printing and to ignore the meaning of the word. Typically, the task is quite difficult and the incongruent trials cause a significant delay in reaction times.

Let us think about a possible frame (Fig. 7) describing the situation in the same terms that have been spelled out above:

Even in this case, there is a graphemic representation of the *English label* that can be included in the mental representation. Being a graphemic representation, it is perceived by the viewer; hence, it makes sense to include perceivable attributes in the frame. The font will have a size and a color, for instance; only the latter is then relevant for the task at hand, which is the individuation of the color. The label that is represented on paper, however, also has a clear connection with a color concept, that includes a portion of color space (and therefore has determined attributes). Now, what can happen in such a representation is that the two portions of color space involved have different values in terms of *saturation*, *brightness* and *hue* i.e. that they identify a different color, possibly named differently. The mental representation becomes, in this sense, more complex and can therefore be the reason why processing costs actually become higher: having to produce a response based on the label given to a color concept, and being the case that two different labels and two different concepts are evoked and involved, the task becomes difficult to solve. Note that the participant
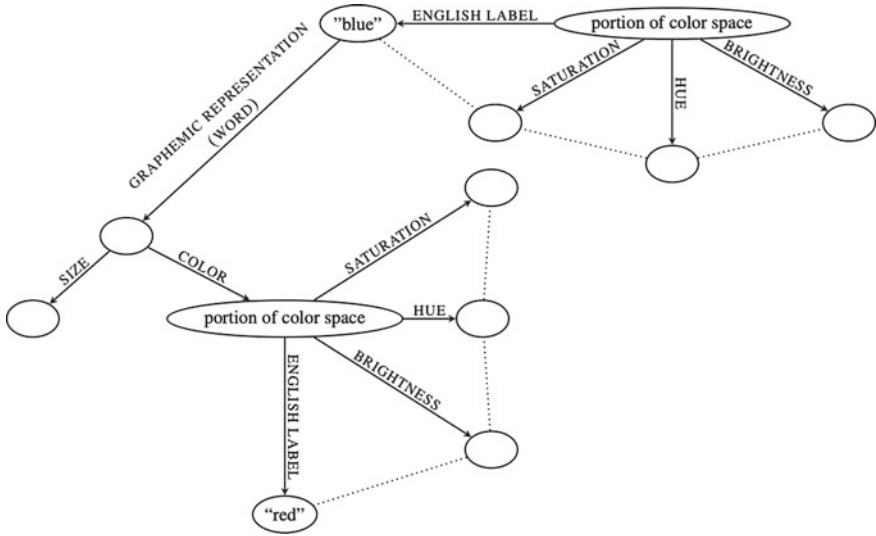
Fig. 7 Frame for a Stroop task (incongruent colors)

does not perceive the label "red" anywhere; however, an attribute is evoked and activated and the task gains complexity and potentially makes it easier to produce mistakes. Having two nodes of the same kind, with the same sort of information, makes it harder to process it, since there is conflicting information regarding the label involved in the task. In a way, this is the opposite of what happens in the case of the blue houses; since the task is not a discrimination one, but rather one where one label has to be produced, the presence of two different nodes of the same kind delays solving the task.

7 Conclusions and Open Questions

In the present paper, a way to model color representations has been proposed that represents them as complex structures used in perception tasks and communicative tasks in a flexible way. The view, as stressed above, is not meant to disprove or support Whorfian-like hypotheses. Rather, the model shows how task requirements shape conceptual retrieval, and how complex representations can be used flexibly in the context of specific tasks in a way that is compatible with the evidence regarding color terms and perceptual tasks presented. Lalumera's suggestion, to consider the idea that "shallow" effects of language labels on non linguistic tasks are still interesting if one does *not* assume mental representations to be rigid units, is here accepted and pushed a bit further: it has been argued that what the evidence suggests is, as a matter of fact, that a view of mental representations that integrates several kinds of information,

recruited flexibly and task-dependently, is indeed able to potentially account for the findings. This idea is implemented in terms of functional attributes representing linguistic information. This is embedded in a view where mental representations are modeled in terms of different kinds of information as functionally integrated in a complex structure, which is what results like that of Pecher and Boot actively seems to suggest and what can be potentially modeled in the Stroop task case.

The presented evidence clearly only gives some clues about how determined mental processes are affected by linguistic labels for perceptual information and about how this can be modeled. The limited set of examples, moreover, can only partially be considered decisive, and the advanced proposal has to be integrated in a full blown theory of frames. The ultimate goal of such a proposal, moreover, would be to have a empirical paradigm that addresses the specific hypothesis regarding the structures of the representations involved. However, the fact that the model seems to be potentially able to accommodate evidence from different research fields is encouraging as far as the possibility to have a better understanding of how perceptual and linguistic information interaction in complex mental representations goes

References

Barsalou, L.W. (1992). Frames, Concepts and Conceptual Fields. In: A. Lehrer, E.F. Kittay (Eds.), *Frames, fields and contrasts*. Hove and London: Erlbaum, Hillsdale, 2174.

Barsalou, L.W., Santos, A., et al. (2008). Language and simulation in conceptual processing. *Symbols and Embodiment: Debates on meaning and cognition*, Oxford University Press. Chap. 13.

Berlin, B, & Kay, P. (1969). *Basic color terms: Their universality and evolution*. University of California Press.

Carruthers, P. (1996). *Language*. Thought and Consciousssness: Cambridge University Press.

Casasanto, D., Lupyan, G. (2019). All concepts are ad hoc concepts. In E. Margolis, & Laurence S. (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 543–566). Cambridge: MIT Press.

Dehaene, S., Bossini, S., & Giraux, P. (2019). The mental representation of parity and number magnitude. *Journal of Experimental Psychology*, 122.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. (2007). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and Language*.

Lalumera, E. (2014). Whorfian effects in color perception: Deep or shallow? *Baltic International Yearbook of Cognition, Logic and Communication*, *9*, 1.

Löbner, S. (2015). Functional concepts and frames. In T. Gamerschlag, D. Gerland, R. Osswald, & W. Petersen (Eds.), *Meaning, frames, and conceptual representation. Studies in language and cognition 2* (pp. 13–42). Düsseldorf University Press.

Macpherson, F. (2012). Cognitive penetration of colour experience: Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research*, LXXXIV(1).

Mahon, Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology*, *102*, 59–70.

Newen, A. (2011). Phenomenal concepts and mental files: Phenomenal concepts are theory-based. *Philosophia naturalis*, *47*(48), 155–183.

Newen, A., & Vetter, P. (2017). Why cognitive penetration of our perceptual experience is still the most plausible account. *Consciousness and Cognition*, *26–37*.

Pecher, D., & Boot, I. (2019). Numbers in space: Differences between concrete and abstract situations. *Frontiers in Psychology*, 2.

Rayo, A. (2011). A plea for semantic localism. *Nous*, *47*(4), 647–679.

Redmann, A., FitzPatrick, I., Hellwig, F., & Indefrey, P. (2014). The use of conceptual com-ponents in language production: An ERP study. *Frontiers in Psychology*, *5*, 363.

Roberson, D., Pak, H., & Hanley, R. (2008). Categorical Perception of color in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition*, *107*, 752–762.

Sapir, E. (1983). In *Selected Writings of Edward Sapir in Language, Culture, and Personality*. University of California Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2006). Russian blues reveal effects of language on color discrimination. *PNAS*, *104*, 7780–7785.

Wolff, P., & Holmes, K. (2010). Linguistic relativity. In *Cognitive Science* (Vol. 2, pp. 253–265). Cambridge: Cambridge University Press.