Speaking for Thinking:

"Thinking for Speaking" reconsidered

Agustín Vicente

Ikerbasque Research Professor,

Ikerbasque/ University of the Basque Country, UPV-EHU

agustin.vicente@ehu.eus

ABSTRACT

Two connected questions that arise for anyone interested in inner speech are whether we tell ourselves something that we have already thought; and, if so, why we would tell ourselves something that we have already thought. In this contribution I focus on the first question, which is about the nature and the production of inner speech. While it is usually assumed that the content of what we tell ourselves is exactly the content of a non-linguistic thought, I argue that there can be a lot of transformation in the process of converting a thought into words. Thus, the content of what we tell ourselves, being intrinsically linguistic, is different from the content of the thought our speech transmits. Fleshing out this kind of approach implies dealing with complicated questions which we lack enough knowledge about: the nature of non-linguistic thinking is, and how speech (inner and overt) is produced; i.e. how the speaker goes from format *a* (format of thought) to format *b* (language). I show that these are pressing issues for any other position, but also suggest ways in which we could tackle such complicated issues.

Keywords: inner speech; thinking for speaking; speech production; linguistic relativity,

Introduction: on what do we when we talk to ourselves

Two connected questions that arise for anyone interested in inner speech are whether we tell ourselves something that we have already thought; and, if so, why we would tell ourselves something that we have already thought. The first question is about the relation between inner speech and thinking. In principle, it is possible to hold either views: that inner speech produces new thoughts, or that inner speech merely expresses pre-existing thoughts. Christopher Gauker (2018) is an example of the first kind of view, while Fodor (2008), who holds that speech in general merely expresses thought, would be a clear case of the second kind of approach. The second question is sometimes voiced as a concern for those who hold that inner speech expresses thoughts one has had. If it is assumed that language is for communication, why would we communicate to ourselves thoughts that we have already had?

In this contribution I do not deal with this second issue (i.e., why we would tell ourselves something we have thought). A plausible, and popular (Vygotsky, 1987) response to it is that by speaking to ourselves we become aware of pieces of information

that help us navigate the world in a controlled way. My focus is the antecedent issue of whether, in fact, we tell ourselves thoughts we have had. This question is not, as the other one, about the function(s) of inner speech. Rather, it is about the nature and the production of inner speech.

In Vicente & Jorba (2019), we explore the possibility that inner speech expresses thoughts that are unlike the thoughts that give rise to inner speech utterances. The process of speaking may involve a reformatting, or representational re-description, that results in a different content being expressed. In particular, there may be two systems of thinking: one that uses a possibly universal format, and another one that uses a linguistic format provided by the particular language that the subject uses. By putting thoughts into words we "translate" from one system to the other by the very act of producing speech. However, such a translation is suboptimal, since both representational systems are quite different (see Section 2). In Vicente & Jorba (2019), we even consider the possibility that the system of "thought" is not propositional. Here that issue is left aside, but I will elaborate on, and motivate on, that view presented in that paper.

Now, the usual approach to the relation between inner speech and thinking is not the one I want to motivate in what follows. Usually, it is assumed that the content of what we tell ourselves is exactly the content of a non-linguistic thought. If one thinks that language merely expresses thought, then it is very natural to think that the content of inner speech utterances is simply mirroring the content of the thoughts that they transmit. To repeat, the view here presented is that there can be a lot of transformation in the process of converting a thought into words, so that the content of what we tell ourselves, being intrinsically linguistic, is different from the content of the thought our speech transmits.

Fleshing out this kind of approach implies dealing with complicated questions about which we lack enough knowledge. The first, most obvious one, is what non-linguistic thinking is like, if it is different from linguistic (inner speech) thinking. A second pressing issue concerns how speech (inner and overt) is produced; i.e. how the speaker goes from format a (format of thought) to format b (language). Thirdly, given what is standardly assumed in the speech production literature, something has to be said concerning the assumption made in that literature since Levelt (1989) to the effect that there is a level of non-linguistic thought whose content is the content of what the speaker says.

Section 1 motivates the general idea of two systems of representation: although the idea of there being two systems of conceptual representation is mostly alien to many philosophers and psychologists, since Slobin (1987) it is implicitly or explicitly assumed by authors working on linguistic relativity or in speech production. This is the idea that there is a level of representation adjusted to the expressive demands and constraints of each language. Slobin called it "thinking for speaking". Section 2 is devoted to explaining a deep concern about Slobin's proposal, namely, that is unclear how these two representational systems interact, or even if they do. Section 3 begins to show that the role that thinking for speaking allegedly plays can in principle be played by speaking *simpliciter*. On this view, one of the two representational systems postulated by several authors would be the language the subject uses, which would be

recruited in the process of producing speech. <u>Section 4</u> compares the view advanced in the previous section against current models of speech production that take it that before speech production begins, conceptual representations are transformed into semantic representations. After Levelt (1989), current models of speech production assume that speech production begins with a "message" that is pre-formatted according to the demands and constraints of each particular language. This is where "thinking for speaking" comes in in these models: conceptual representations are "translated" into the semantic representations that constitute the "message". However, this is also where the interaction problem explained in Section 2 kicks in. I suggest that doing without this intermediate station, the "message" level as it is conceived nowadays, is overall better. We should consider that speakers go directly from perception or conceptual thinking to language, and that in so doing they create new thoughts. In the last section, "summary and conclusions", I compare this view with other views about the nature of inner speech, with a particular focus on the Vygotskyan approach.

1. Two representational systems: CR and SR

Linguistic Variation

It is often said that human beings think with representations that have two characteristics: first, they are conceptual; and second, they enter complexes that are propositional. It is often thought that the second characteristic follows from the first, since concepts are often defined as the components of thoughts, that is, propositions. However, in principle, a creature could have concepts but not have propositional representations (although not according to criteria such as the famous "generality constraint" of Evans, 1982). For quite some time, it has been assumed that the representations with which we think are representations of a language of thought or Mentalese (Fodor, 1975, 2008). However, more and more authors are skeptical about two issues: (i) that the only representations with which we think are the representations of Mentalese; and (ii) that we actually think in Mentalese. This section focuses on the first question, and specifically on the evaluation of the widespread idea that there is a level of conceptual representation that co-varies with linguistic variability. This is a proposal that is gaining traction lately. According to it, there may be two representational systems that we think with, or at any rate, our mind works with: the first one uses conceptual representations that are likely to be universal, and the second one uses conceptual representations that are aligned with the semantic representations of the language of each individual.

Following Levinson (2003), we can call the first type of representations "CR" (for "conceptual representations"), and the representations of the second type, "SR" (for "semantic representations"). It is considered that CR and SR have to be different representational systems because each language makes partitions of the different realms of reality in different ways (this is the "linguistic diversity" thesis). For example, if we look at the domain of spatial relationships of inclusion and support, we see an interesting variation just by looking at a few languages. In Spanish there is a generic *en* that applies to the two types of relationship, inclusion and support. In English it is mandatory to distinguish between both relationships by means of *in* and *on*. Dutch distinguishes at least three support relationships: *op* for "robust" support (eg, a cookie

on a plate), *aan* for "tenuous" support (eg, clothes on a string) and *om* for the relationship from "surrounding" (eg, a necklace around the neck). Finally, in Korean there is a tripartite distinction: *kkita* denotes both supportive and inclusive relationships that are tight (a Lego piece fitted into another, or a ring on a finger), *nohta* refers to loose inclusion relationships (a pen in a drawer), and *nehta* applies to equally loose support relationships (a cup on a table).

Examples of cross-classifications or simply mismatched linguistic partitions abound. The domain of color has been particularly and extensively studied (Davidoff, et al., 1999; Kay et al., 1997; Regier & Kay, 2009). The domain of motion events has also received a lot of attention since Talmy (1985) distinguished two fundamental types of languages: those that encode manner of motion (for example, English) and those that focus on trajectories (for example, Spanish). In general, an English speaker will describe a scene in which an individual runs from point *a* to point *b* by making reference to the manner in which they move (*run*), while for Spanish speakers, referring to the manner of motion is expensive and anomalous. Their description will skip manner of motion but be more accurate about the trajectory (*S* went from *a* to *b*). The result is that English speakers might apply the same description to two motion events that Spanish speakers might describe in different ways, and vice versa.

In recent times, authors such as Barbara Malt (Malt et al., 2011, 2015) and Asifa Majid (Majid et al., 2008) have done very interesting work on cross-linguistic differences and similarities in domains such as the differences between *breaking* and *cutting* (Majid et al., 2008) or human locomotion movements (*walking*, *running*, *jumping*, etc., Malt et al., 2008). All these studies show that there is great variability in how the different languages partition the different domains of reality, despite the fact that the work of these researchers also reveals coincidences at an abstract level[1].

CR and SR: roles in cognition

Taking into account such differences in the classifications that each language makes of the world, several authors conclude that Fodor's idea that the semantics of English is the semantics of Mentalese (Egan, 2010) is untenable. Whatever representational system we think in, its semantics has to be importantly different from the semantics of the different natural languages. More importantly, the linguistic variation data (together with other assumptions) suggests that there has to be a level of mental representations whose semantics is aligned with the semantics of each subject's language.

Positions concerning the cognitive roles of CR and SR differ. According to Levinson (2003), who first introduced the distinction in the way we have put it, both CR and SR are representational systems that are used for thinking (or, as he puts it, for "serious thinking"). However, for Slobin (1996) and Papafragou (e.g. Papafragou & Grigoroglou, 2019), the CR system is the system of thought, while SR is only used in the process that Slobin calls "*thinking for speaking*", which implies reorganizing the thought that the speaker wants to communicate so that it is easily articulated in his own

---

[1] For example, a difference between cutting verbs and verbs of breaking that is respected almost universally revolves around the predictability of the result of the destruction process of the object: it is predictable in cutting, but not in breaking.

language. Basically: when subjects want to communicate a thought, they restructure such thought, so that it is easier to put it into words.

It is not clear, yet, when (or how) such a process occurs. Following some of Levelt's remarks (1989), Papafragou limits "thinking for speaking" to the process of formulating a message as part of effective linguistic production. Slobin (1996) is more liberal: subjects may be adapting their thoughts to communicative exchanges that might occur, although in fact they do not occur. Thus, it could happen that the use of SR is very widespread, since it could happen that we are often preparing our thinking for the eventuality of having to communicate it at one time or another. The construct "thinking for speaking", therefore, is quite vague and it is difficult to know how to operationalize and measure it.

On the (alleged) indispensability of both CR and SR

A reason for postulating two different types of representational systems for thought relates to the appearance of neo-Whorfian hypotheses, even in their mildest forms (Gumperz & Levinson, 1996). As mentioned, Levinson, a strong neo-Whorfian, considers SR to be the main representation system we use in thinking, while Slobin restricts it to "thinking for speaking". If relativistic effects (i.e. effects of language on cognition) are to be explained, then it seems that something analogous to SR has to be postulated. Yet, not only neo-Whorfians believe that SR are indispensable. For instance, Papafragou, otherwise anti-Whorfian (see Gleitman & Papafragou, 2005), argues that SR play an indispensable role in preparing the message for speech production. According to her, Levelt (1989) himself already indicated that the type of representation with which the content to be communicated is formed (the "message", the first level in the sequence of speech production) has to conform to the expressive requirements and preferences of the speaker's language. This idea has become a widely used assumption in speech production research (Papafragou & Grigoroglou, 2019).

On the other hand, it does not seem easy to be eliminativist about CR. If we do without CR, a universal conceptual system, we are doomed to a strong whorfianism whose evidential support is currently non-existent and which has well known problems. To put some classic examples from Arnauld and Nicole (1996) in their *Logic or the Art of Thinking*: (i) if we could only think according to the way our different languages articulate the world, it would be impossible for a French speaker to understand how a German speaker behaves and vice versa; and (ii) explaining how linguistic conventions arose requires that they have some prior way of thinking about things in the world.

So, we can suppose that there is a CR representational system due to the problems that would result from getting rid of such an assumption, even if, as of today, we barely know anything about CR. We know that pre-linguistic babies have some concepts, but we can even doubt that they are concepts *tout court*. According to Carey (2009), the concepts that make up the so-called "core cognition" of pre-verbal babies are iconic and are not articulated in propositions. It is difficult to think that such concepts make up adult-like CR. On the other hand, the study of infant cognition seems to give us clues about what CR could consist in and even what its basic functionality could be. A nice example is Hespos and Spelke's (2003) work on spatial relations in infants.

Hespos and Spelke (2003) show that the Korean system of spatial relationships (see above) coincides with infants' categorization of spatial relations. Moreover, they hypothesize that this is due to infants' interest in predicting motion patterns: two objects in tight support or inclusion move together; an object loosely contained in another moves with it, but with some degree of freedom; finally, an object that is simply on another (a glass on a table) does not have to move with this second object. It is quite likely that this type of categorization persists in linguistic adults.

2. The translation problem

The problem

Now, with both CR and SR in place, it is time to rise a crucial issue that has not been examined. Suppose subjects have thoughts in CR that they want to express linguistically at some point: how do they go about translating them into the SR system? The mapping from CR to SR is a *many to many* mapping. For example, suppose that the CR thinking about spatial relations of inclusion and support does indeed reflect the Korean system (Section 1). A "tight" relationship can be translated as *on* or *in*, depending on the case. On the other hand, an *in* relationship can be tight or loose.

Translating from one representational system to another and back when the mapping is many-to-many, is always a complication. There may be well-established habits of translating some representation units into other representation units, but since both systems are assumed to be systems of discrete infinity, the process of translating a complex representation in the SR system to a complex representation in the system CR (or vice versa) cannot be made automatic. Translation possibilities increase with the length of the representational complex. Furthermore, the mapping between CR and SR representations becomes more complicated by virtue of the different expressive powers and demands of the two representational systems.

Concerning expressive powers, SR units can correspond to CR complexes, as well as vice versa: *tight-fit*, a CR unit that represents the tight fit relation for both inclusion and support, translates into two possible SR noun + adj. compounds: *tight support* and *tight inclusion*. SR representations, on the other hand, make some CR representations difficult to express (for example, expressing manner of motion is expensive in Spanish).

Concerning expressive demands, it is at least plausible to think that SR include information that may be irrelevant from the point of view of the conceptual system. For example, the evidential system in some languages such as Turkish forces subjects to express where the information that is transmitted comes from. On many occasions, this type of information will not be relevant for the cognitive processes of the subject, which means that probably CR will not reflect it.

So, the translation from one system to the other is a necessarily complicated affair. Moreover, how do subjects know when they have to go from one thought system to the other? If the translation is triggered by the need to speak, for example, then there is an answer to this last question, but if the translation is triggered by something like "preparing to verbally report thoughts", it is unclear under what conditions the CR-thoughts will be reformatted according to the requirements of SR. The beginning of the translation process would not be under the control of the subject, but neither would it be

under the control of some specifiable external triggers. It is even more complicated to explain when we would go from CR to SR and vice versa if both systems of representation are alternative systems of thought (Levinson, 2003).

In general, it would be preferable to have a theory that simplifies both the process of translation from one representational system to the other, as well as the question of what triggers such process. We must also ensure that these two representational systems are not redundant. That is, taking into account the little we know about each of the systems, plus the translation problem that arises, we have to make sure that each system has a specific role that the other system cannot fulfill. It will be argued later on that the process of "thinking for speaking" is an unnecessary complication. This implies that SR do not play any cognitive role other than the cognitive roles that speech production can play (in the form of inner speech or in unconscious labeling: see below).

Independent systems?

One possible answer to the translation problem is to hold that CR and SR never interact. For example, depending on what subjects have to do, they could recruit either CR or SR. If they have to think for speaking, or for any other role that we allegedly use SR for, they will recruit SR. If their thinking is not related in any way to communication, they will use CR. A drastic way to avoid interaction between the two systems is to put them in different places. The "Whorf was half right" (WHR) hypothesis, advanced by Regier and colleagues (Regier et al., 2010; Roberson et al., 2008) could support such a view. According to this hypothesis, with ample empirical evidence to support it, the left hemisphere categorizes the world according to the subject's language, while the right hemisphere categorizes the world according to universal representations. As has been said, there is ample evidence supporting this hypothesis: for example, it has been seen that, in a color discrimination task, linguistic categorical perception (i.e. categorical perception aligned with the lexical distinctions made by each language) appears only if the stimuli are projected in the right visual field.

In these tasks, subjects are asked to identify a sample of a color that slightly differs from the color of other samples that together form a circle. If subjects speak languages that distinguish the color of the odd sample from the color of the rest of the circle (say, *green* vs *blue*), reaction times are faster than if subjects speak languages that do not mark the distinction between the odd sample and the rest. The experiments by Regier and colleagues show that this type of categorical perception occurs only in the right visual field, which is projected onto the left hemisphere, the "linguistic" hemisphere. Interestingly, under verbal interference, such a categorical perception in the left hemisphere disappears. According to Regier and colleagues, we can see these types of effects in different domains, not only in the domain of color. Thierry and his colleagues (Thierry et al., 2009) have also found this type of lateralization: conceptualizing a visual input or sound associated with some entity *F* activates areas in the right hemisphere, while interpreting a word that refers to *F* activates areas in the left hemisphere. However, the results of Thierry and his colleagues (2009) are more nuanced, since while they find lateralization in a first stage, they also observe that either conceptualization path ends up activating common areas in subsequent processing.

It is true that there are "revision" effects that may seem problematic for the hypothesis that the two representation systems belong to processing systems located in different areas of the brain which do not interact. For example, Li et al. (2009) showed that the different ways of classifying triads of objects (either by form or by substance, depending on whether or not the languages have count terms, as shown by Lucy and Gaskins, 1996), can be easily reversed simply by making subjects aware of the alternative way of classifying things. The Lucy and Gaskins experiments (see also, Lucy, 1992), showed that speakers of languages without count terms classified more by substance (e.g. being made of cardboard) than by shape, while language speakers with count terms tended to group objects by similarity in shape. Li et al. (2009) presented the subjects with a set of items of different substances and shape so that they classified them according to their similarity. Then they were asked about each of the items, if that item was an object or a substance. They replicated the results of Lucy and Gaskins in the first phase of the experiment (i.e. speakers of different languages tended to classify the items, some according to shape similarity, while others according to shared substance). However, the responses to their question were uniform: the speakers of the two different types of languages are not blind to the distinction between objects and substances; that is, they can group the items also according to supposed universal CR.

Prima facie, these types of results show that the supposed use of SR in categorization tasks can be "corrected" by using CR. On the other hand, the work of Landau et al. (2010) that is discussed below shows that linguistic representations can have an effect on how subjects categorize and, therefore, that it is also possible to "correct" categorizations made without language. What Landau and his collaborators argue is that linguistic stimuli improve performance in some tasks, such as remembering the spatial distribution of colors in a scene, a task that involves attaching one type of properties (colors) to others (space regions). In this sense, if what linguistic stimuli do is to activate SR representations, then it seems that the use of CR can also be "corrected" (in this case improved) by using SR representations. However, this kind of corrections that work in both directions do not imply that the CR and SR systems actually interact. What the studies show is that subjects can use one system or another even when they have previously used the alternative representational system. However, the main concern for the general SR / CR distinction derived from the work of Li and colleagues (2009) and of Landau and colleagues (2010), among others, is the interpretation they make of their findings. Interestingly, such an interpretation fits well with the findings of Regier et al. (2010). The first part of the next section discusses how such an interpretation puts pressure on the idea that there is a SR level. The discussion of interaction issue resumes in Section 4.

3. The indispensability of semantic representations revisited

Offline effects and online effects of language

The evidence amassed by Regier and his colleagues does not suggest that we use a system of semantic representations (SR) in the left hemisphere. The evidence suggests a schism, but not of different conceptual systems, CR and SR, but of different categorization systems: verbal and non-verbal. We have mentioned before the notion of offline *effects* of language on thought. Offline effects are distinguished from *online*

effects: language has offline effects on thought if it has effects on cognition regardless of whether the linguistic system is active, for example, if it has effects on what Whorf (1956) called "habitual thinking ". Language has online effects in cognition if the effects are caused by the activation of linguistic representations: lemmas, lexemes, phonological representations, etc., that is, representations that are part of the speech production process. The fact that the Whorfian effects reported by Regier at al. (2010) disappear under verbal interference seems to show that such effects are online, that is, they are the result of producing linguistic labels.

According to Li et al. (2009), none of the relativistic effects found so far definitely shows that there are offline effects of language, or that we use the SR system in our thinking. The possibility that the subjects are using language in one way or another when performing the experimental tasks, according to these critics, has not been ruled out, and is in fact the most likely explanation for the differences observed in the categorization tasks. Landau et al. (2010), on the other hand, argue that their findings on how language can enrich the way we represent a scene (so that we can later remember more details about it) are better interpreted if we think of language as a means to focus attention. The linguistic stimulus does not activate a special type of representations, but helps subjects focus their attention on aspects of the scene that they would not have encoded.

The visual system has well known problems about binding features to a visual representation. For example, we can see a square divided into two halves by a vertical line, the right part being green and the left part red. When, after a brief exposition, we have to select the square that we have seen among four images with different distributions of green and red, we exhibit problems about retrieving the correct distribution of colors. We know that the colors in question are green and red, but we do not know where each color was located: we do not bind the colors to the correct left and right parts of the square. Landau et al. (2010) tested whether binding colors to spatial regions was improved through language instruction. They tested four-year-olds under different conditions: (i) the experimenters labeled the entire square ("*this is a dax*"); (ii) they said "*the red is touching the green*"; (iii) they called children's attention to the red part ("*look at the red!*"); and (iv) they directly said: "*the red is on the left*". Children improved over the no-linguistic condition only in the last of these conditions, even if they did not master the right/left distinction.

According to Landau et al. (2010), the results do not suggest that the children retained only the linguistic representation and discarded the visual one, but rather that the linguistic stimulus made them direct their attention towards the spatial relation between the two colored regions, thus simplifying the task of binding properties of a different nature. The effect of the linguistic stimulus, however, is temporary: after ten minutes, children were not able to distinguish the original square from its mirror image.

In general, the evidence seems to show that there are many online relativistic effects, resulting from the use of linguistic labeling (Lupyan, 2012; Tan et al., 2008). Such labeling begins early in human development, to the point that it is already present at two years of age (Kahn, 2013). Labeling implies that subjects recognize that the stimulus is nameable in a certain way. However, categorical perception, and with it the relativistic

effect, seems to emerge in the moment when the linguistic label is used. Otherwise, the effect would not disappear under verbal interference.

Enriched cognition

However, not all relativistic effects can be easily explained as online labeling effects (or, as Li et al., 2009, put it, as language-on-language effects). A case in which relativistic effects do not seem to be due to the effective production of language is effects related to the use of spatial metaphors to refer to time. Both Casasanto and Boroditsky in several articles (e.g., Casasanto, 2008; Casasanto & Boroditsky, 2008) have shown that subjects' conception of time varies with the way in which their language draws the analogy from space to time. For example, if a language frames time in terms of quantity, speakers of said language seem to conceive of the passage of time as an accumulation of a certain quantity. If, in contrast, the language frames time in terms of distance, speakers of said language will tend to conceive the passage of time in terms of movement along a path (Casasanto, 2008).

It is important to note that the tasks used to test these effects do not involve linguistic production. For example, Casasanto's experiments evaluate whether speakers of languages that use the metaphor time-as-length ("long time") judge that, just by virtue of a line growing longer than another, the former line's growing took longer than the latter's. Alternatively, he tests if speakers of languages with the time-as-quantity metaphor ("much time"), judge of a container *a* that gets more volume of a liquid than another container *b*, if it took longer to fill *a* than to fill *b*, even though in actuality it takes the same exact time to fill both containers. This kind of experiments suggests that, depending on which language we use more, we represent the time flow either as a moving line or as an accumulating quantity. This is proof that there are offline effects of language.

Yet, the main issue should not be so much whether language can influence the way we think, but whether such effects give us reason to believe in a SR system. If offline effects were massive, we would have to conclude that we do think in SR, after all. However, research has mostly found offline effects in localized and quite abstract domains, which suggests another hypothesis, that of "enriched cognition"." It may be that we enrich our CR system with concepts provided by our language in domains that are difficult to think about without language. The idea of "enriched cognition" was endorsed by Fodor (1975). According to him, representations of categories provided by our different languages could be incorporated into Mentalese by a process of "chunking" (i.e., forming a new representational atom by grouping into a single representational unit the content of several atomic representations). However, incorporating linguistic representations into Mentalese would not alter the fact that we think in Mentalese. Similarly, in the case of metaphorical conceptualizations of time, subjects could be incorporating new representations to CR, but such possibility does not imply that such subjects think in SR, i.e., a different level/system of representations. In short, the case of metaphors suggests that the incorporation of a metaphor made conventional by a certain language facilitates thinking about a certain domain, but not that there is a whole system of mental representations that corresponds to the categories represented by the different languages. There is no reason why only metaphorical

extensions made conventional by a certain language should have this kind of impact. However, there is also no reason, in principle, to think that more offline effects of language could not be accommodated by the hypothesis of enriched cognition. Only the discovery of massive offline effects would constitute evidence in favor of an independent SR system of thought.

The discussion concerning offline and online effects of language is intended to put pressure on the existence of a SR system in thought. The WHR hypothesis above was introduced as a possible development of the idea that SR and CR do not interact, and, therefore, as a possible way to avoid translation problems. However, the evidence supporting the WHR hypothesis does not support the idea that there is a SR level, after all. If Whorfian effects are online, SR does not have any explanatory role to play in that area. On the other hand, the eventual problem of translation is effectively dissolved if linguistic categorization and non-linguistic categorization occur in different parts of the brain. However, it is important to note that the no-translation situation only would happen when subjects are confronted with some perceptual stimulus. In such a case, which is the one most experiments investigate, subjects may categorize the stimulus either using their language or using other conceptual resources. But, of course, there are many situations where we produce language in the absence of a perceptual stimulus. In such situations, we have to go from thought to language. That is, it is unlikely that thought and language do not interact.

The "message" in speech production

While the previous subsections have dealt with the (in)dispensability of SR in thought, this subsection focuses on the (in)dispensability of SR on speech production. In research on speech production, it is typically assumed that the message to be transmitted has to come already prepared ready for emission. That is, the message that a speaker is trying to communicate has to be represented in a way that the linguistic processor can easily "understand". Since Levelt (1989), many authors have assumed that the linguistic production system requires an input preformatted according to the requirements imposed by the particular language the speaker speaks (see Papafragou & Grigoroglou, 2019). Thus, if the speaker is Spanish-speaking, the message to be expressed regarding a motion event will represent trajectory, but not necessarily manner of motion, while if the speaker is English-speaking, the message will represent manner of motion. Likewise, if the speaker speaks Korean and wants to describe a spatial relationship between two objects, its message will contain information about the tight/loose dimension, information that will not be part of the message of a Spanish speaker. If the speaker speaks English and wants to report on a past event, she may be indifferent regarding the temporal aspect –perfect or imperfect-, but not if she speaks in Spanish. If the speaker speaks German, she will not mind distinguishing progressive from usual, but if she speaks English, she will, etc.

Therefore, according to this idea, at least the message to be communicated comes in SR. The process of "thinking for speaking" and, therefore, of conversion from CR to SR. has to take place at least in the moment of encoding the message to be expressed (that is, when subjects are about to speak).

However, the benefit of having the message prepared in this way is not obvious. Having the message formatted according to linguistic demands and constraints makes it easier to move from the thought to the language that will express it. However, postulating a message level only moves the problem of translation one step upwards. The thought that the speaker wants to express is a CR thought at some point; then, it is translated to SR, and finally it is expressed in words. Perhaps the last step, moving from SR to words, is easy, although polysemy and synonymy still create a many-to-many mapping problem, as Bierwish and Schreuder (1992) showed long ago: the content of a certain SR can be expressed with different linguistic expressions (synomy), while different SR contents can be expressed using the same linguistic expression (polysemy). However, the first step, going from CR to SR is not easy at all, as it has been shown above.

What evidence is there for this kind of "message" level? According to Papafragou & Grigoroglou (2019), for example, there is sufficient evidence that messages (but not thinking in general) uses the SR system. Evidence comes from a good number of eye-tracking experiments that Papafragou and colleagues, among others, have conducted comparing English speakers against, for example, Greek speakers on motion events (e.g., Papafragou et al., 2008). In one condition, subjects simply inspect a scene that represents a motion event. In the other condition, subjects are told that they will have to describe the event at some later time. In the second condition, Greek speakers spend more time inspecting the trajectory elements than do English speakers, while in the first condition (free inspection) no significant differences are found between English and Greek speakers.

According to Papafragou (2015), these results imply that in the free inspection condition, subjects conceptualize the motion event in a similar way, while in the later described condition, subjects engage in "thinking for speaking", i.e., preparing a message that meets the demands imposed by the subject's language so that it can be easily put into words. However, this kind of experiments does not rule out that subjects in the later described condition are in fact producing language. On the face of it, a natural way to code an event that needs to be described later is to actually describe it.

In order to account for some intriguing effects in some studies of the free inspection condition, Papafragou (2015) suggests that subjects may be using language to encode certain information. The effects in question refer to final inspections in dynamic scenes that represent motion events. They observed that the subjects who were instructed to inspect and remember such events took some final glances, either at the goal (Greek-speaking subjects), or at the moving object or person (English-speaking subjects). However, under verbal interference, such an effect disappeared, suggesting that, after encoding the event in CR, they went back to encoding relevant information, making active use of language in order to consolidate their memories. This seems to imply that there is no direct evidence that in the describe-later condition, subjects are not already producing (inner) speech.

4. Speech production without the message level

Speech production without SR: from perception to language

There are two main views concerning message encoding in speech production. One is the holistic view, according to which the message is fully encoded before speech production begins. The other model is the incremental one, according to which the message is encoded incrementally, and in parallel to the execution of "subordinate" processes such as lexical access, the formulation of the syntax, etc. In principle, it seems that only if speech production were incremental and parallel there would be room for a model that dispenses with the message level. Spanish speakers could be observing a scene and retrieving words that partially describe it incrementally: for example, they could retrieve the word *salir* ("leave") to describe a scene where someone drives from point A to point B. The word will be inserted into a tree structure, with a slot for an external argument (someone does leave) and a slot for a locative prepositional phrase (someone is leaving from somewhere). The construction of such tree structure would force the speaker to search for words to describe the agent and the place where the agent is going, as well as the morphemes necessary to express time and aspect. Spanish speakers could also add an adjunct PP indicating destination. In contrast, English speakers could begin by labelling the motion event with the word *drive*, expressing manner of motion. Then it is optional to them to also and express the origin and end of the trajectory in the form of adjunct prepositional phrases (*drove from A to B*).

Therefore, incremental message preparation could be just incremental production, based on acts of linguistic labeling without prior conceptualization (labeling itself constitutes the act of categorization). However, it seems that it will be more difficult to dispense with the preparation of a previous message if the supposed process of preparing the message is holistic. If there is evidence that speakers wait to gather all the relevant information before they start looking for the words to express it, then it seems clear that they move from perception to full conceptualization, and from there to speech production.

However, the holistic model does not have to commit to this particular reconstruction of the process. Holistic message planning could also consist of building a tree structure in a less obvious incremental way. Instead of starting to build the tree with a label, and then going back to check what other words to insert, adding branches, etc., speakers may store words in working memory and build the tree only when they know they can express all the relevant information. Since words contain syntactic information, building a tree from a list of words can proceed relatively smoothly.

In any case, the type of experiments that are used to investigate message encoding, which generally have subjects describe scenes, can in principle be explained from a straightforward linguistic encoding model, which dispenses with message encoding entirely. This is what we also found when the empirical evidence concerning the effects of language on categorical perception was revised.

Speech production without SR: from thought to language

There is another situation in which we speak, namely when we are not talking about a scene we are experiencing. In such case, there has to be a translation from CR to the representations of the language of the speaker. In the model proposed here, going from CR to language does not require recruiting SR, which would only act as intermediate relay stations.

For example, in the experiments of Papafragou et al. (2008) speakers of different languages who are instructed to simply inspect and recall the scene do not differ in their eye fixation patterns. This suggests that they are encoding the information in the same way, presumably in CR. When these subjects want to express what they have conceptualized, how do they do it? This kind of situation has not been investigated for obvious reasons, as it is not easy to know how to recruit the experimental evidence. Research in speech production has focused on the kind of controllable case in which speakers have in front of them a scene they have to describe. That is, the task is near and is the same for all subjects, apart from the fact that the best instrument to study speech production today is the eye-tracker, which requires using a visual stimulus. The same applies to categorization tasks discussed in the first part of Section 3.

Now, in Section 2 we introduced the translation problem affecting CR and SR interactions. From the argument developed up to this point, it follows that the SR representation level is unnecessary: we do not need such a representational system to explain what occurs in thought, nor does it seem necessary to resort to it to explain speech production. So, we could say that the translation problem is "solved". Yet, if we do it without the SR system, the translation problem persists in another form: CR representations have to be turned into linguistic representations anyway, taking into account that the mapping of each unit is many-to-many and the complexity of complex representations is in principle unlimited. The speech production model that we would have to adopt would imply the existence of a message level encoded in CR representations that is then translated into the words and the syntax of a given language.

As said above, we know very little about the CR system to date. In principle, it can be speculated that it is an action and prediction oriented system. For example, if the CR system does indeed incorporate the distinction between tight and loose spatial relations, it is for reasons that have to do with predicting how two objects in loose and tight relations will move. Similarly, the CR system might distinguish between cutting events and breaking events based on the respective final states of both kind of events: controlled/predictable destruction of the integrity of an object (*cut*) vs uncontrolled/unpredictable destruction of the integrity of the object. Thus, from a categorization system possibly oriented to intervention, we would have to move to a representation system, the linguistic one, whose constraints have to do with efficient communicability, constraints that are compatible with great variability. In addition, each language has its own demands about what must be expressed and their own constraints about what can be expressed.

The translation process would of course be easier if it were done incrementally so that the mapping could be done unit by unit. For example, if speakers have to express their conceptualization of an event, they could begin by translating the representation of the kind of event into a verb. Then, based on said verb's argument structure, speakers would have a syntactic tree with gaps for arguments. From that moment on, the translation process would be guided by the expressive needs of the language, which can be converted into instructions like: look for the agent, look for the patient, look for the source of information (in the case of evidential languages), etc.

Alternatively, and as proposed by some authors such as Matchin & Hickok (2019), the speech production process could start with an empty tree that would reflect the structure that the language would give to a given kind of event. Actually, both kinds of approach to the translation process can coexist (Matchin & Hickok, 2019). The important thing is that the process is simplified in the moment in which speakers come up with a syntactic structure that can be used to capture the content of the CR representation of the event. From that moment on the translation process is guided by the demands of the speech production system, i.e. the output system, which effectively reduces degrees of freedom in the translation process.

Quite possibly, tackling the translation problem involves revising the prevailing model of speech production. Since Levelt (1989), it is assumed that, after preparing the message, the first step in production is the recovery of lemmas or representations of words. Next, lemmas are structured in a morpho-syntactic structure. The process does not have to be sequential (Hickok & Poeppel, 2007), but the fact is that the recovery of lemmas is considered a priority over the process of morpho-syntactic "formulation". However, if we dispense with the level of the message as it is usually conceived, the "formulation" process possibly becomes the first step in the production of speech, perhaps after the recovery of some few lemmas. This seems the best way to simplify the process, since it is the most effective way for the output system to impose its constraints and demands.

Note, on the other hand, dispensing with the SR level has a considerable advantage. The process of translating a CR thought into the language of the speaker can be done following the rails provided by the grammar of said language. However, the conversion of a CR thought into an SR message does not have a comparable aid. In other words, skipping the SR level can make the translation problem more tractable.

5. Summary and Conclusion

There are probably two representational systems of thought characteristic of human beings: the universal conceptual system and the linguistic system. That is, we can think with concepts that are developed largely outside of language, but we can also think by producing language. It is doubtful, however, that we use an additional system of conceptual representation, either to think or to speak. Such system seems to introduce complications, rather than to solve them. The consequences of dispensing with the SR system are considerable in the field of psycholinguistics, not so much because it forces to reconsider the prevailing model of speech production, but because it brings to light a translation problem that has not been properly addressed so far. The problem already existed, or exists, but remains hidden because the postulation of SR representations at the message level allows the production researcher to ignore the question of what happens prior to the configuration of such message.

This concerns speech production in general. However, when we move to inner speech, i.e., to cognitive uses of language, dispensing with SR has an unanticipated consequence, namely, that our inner speech utterances have contents that we have not previously thought. In the case of perception to speech, we could be directly thinking in inner speech; for example, having thoughts in inner speech we have not had in any other format. In such case, our inner speech thoughts would be entirely new. In the case of

conceptual representations to speech, moving from unconscious thoughts couched in CR to speech would transform the content of said thoughts considerably, in a manner that our conscious inner speech thoughts would be *about* different things. Inner speech thoughts would be about, for example, support relations, trajectories, the imperfective nature of the past event, the source of information, etc., whereas the thoughts that gave rise to such inner speech thoughts would have been about tight/loose fit relations, manner of motion, (simply) the past , etc.

It remains to be explored what this view would imply. In particular, we should investigate what kind of difference inner speech thinking makes; for example, what is the difference between using inner speech and not using it may be in more or less practical matters. The view suggests that by talking to ourselves we give shape to our thinking. If so, fostering the use of inner speech in education and as interventions would be important not just because of the usual reasons, related to improving self-regulation, but also because the very process of thinking would be intimately related to speaking.

This proposal is congenial to Vygotsky's overall view about inner speech, and in particular to his ideas that inner speech exemplifies the cultural line of development, and that language, or better, speech, plays a major role in shaping thought (Vygotsky, 1978, 1987). It is also congenial to the Vygotskyan picture in that it links verbal thinking to consciousness (and therefore, cognitive control) because, as explained in Martínez-Manrique & Vicente (2015) and Vicente & Jorba (2019), by putting thoughts into words we make the content of our linguistically elaborated thoughts conscious. However, there are also differences between the current proposal and the Vygotskyan approach. Vygostky (1987) insisted that once speech is internalized in the form of inner speech, it is transformed in several ways: inner speech is syntactically condensed and semantically idiosyncratic. The current proposal gives a prominent role to syntax in elaborating linguistic thoughts, which means that syntax has to be fully articulated. However, note that this does not imply that whole sentences have to be pronounced. In speech production models, once syntax is completed, information is sent to the phonological component, and from there to motor commands. So speakers may not "give voice" to all words in their thought as they move from syntax to phonology and motor commands, or they may even refrain from giving voice to their thought at all, giving rise to the phenomenon of "unsymbolized thinking" (Vicente & Martínez-Manrique, 2016; see also Grandchamp et al., 2019). In sum: condensation may appear at the level of pronunciation, but such condensation is not a sign of *syntactic* condensation. Regarding semantics, according to the view here developed, speakers recruit the word-meaning pairs of their own language to describe the world that they perceive or to elaborate the thoughts they have. In principle, there is no difference in this regard between speaking to others and speaking to oneself, which means there is no room for idiosyncratic, personal, associations between sign and meaning. However, these points of disagreement with Vygotsky's view are minor in what I take to be a like-minded approach to inner speech.

Acknowledgements

References

Arnauld, A., & Nicole, P. (1996). *Logic or the Art of Thinking*, translated by Jill Vance Buroker. Cambridge University Press.

Bierwisch, M., & Schreuder, R. (1992). From lexical concepts to lexical items. *Cognition, 42*, 23–60, DOI: 10.1016/0010-0277(92)90039-k.

Carey, S. (2009). *The Origin of Concepts*. Oxford University Press.

Casasanto, D. (2008). Who's afraid of the Big Bad Whorf? Cross-linguistic differences in temporal language and thought. *Language Learning, 58*, 63-79, DOI:10.1111/j.1467-9922.2008.00462.x

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition, 106*, 579-93. DOI: 10.1016/j.cognition.2007.03.004

Davidoff, J., Davies, I.R.L., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature, 398*, 203-204. DOI: 10.1038/18335

Egan, F. (2010). *Mentalese*. In A.Barber & R. Stainton (Eds.) *Concise Encyclopedia of Philosophy of Language and Linguistics*.Elsevier

Evans G. (1982) *The Varieties of Reference*. Clarendon Press.

Fodor, J. (1975). *The Language of Thought*. Harvard University Press.

Fodor, J. (2008). *LOT 2: The Language of Thought Revisited*. Clarendon Press.

Gauker, C. (2018). Inner Speech as the Internalization of Outer Speech. In Peter Langland-Hassan & Agustin Vicente (eds.), *Inner Speech: New Voices*. Oxford University Press. pp. 53-77.

Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. Holyoak and R. Morrison (eds.) *Cambridge Handbook of Thinking and Reasoning*, . Cambridge University Press.

Grandchamp R., Rapin L., Perrone-Bertolotti M., Pichat C., Haldin, C., Cousin E., Lachaux J.P., Dohen M., Perrier P., Garnier M., Baciu M., Lœvenbruck H. (2019). The ConDialInt Model : Condensation, Dialogicality and Intentionality dimensions of inner speech within a hierarchical predictive control framework.

Frontiers in psychology: Exploring the Nature, Content, and Frequency of Intrapersonal Communication, 10. DOI : 10.3389/fpsyg.2019.02019

Gumperz, J., & Levinson, S. (1996). *Rethinking Linguistic Relativity*. Cambridge University Press.

Hespos, S. J., & Spelke E. S. (2004). Conceptual precursors to language. *Nature, 430*, 453-6. DOI: 10.1038/nature02634

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature, 8*, 393–402. DOI: 10.1038/nrn2113

Kay, P., Berlin B., Maffi, L., & Merrifield, W. (1997). Color naming across languages.In C.L. Hardin & L. Maffi (Eds.), *Color categories in language and thought*. Cambridge University Press

Kahn, M. (2013). Thinking in Words: Implicit Verbal Activation in Children and Adults. PhD Thesis. Harvard University.

Landau, B., Dessalegn, B., & Goldberg, A. M. (2010). Language and space: Momentary interactions. In P. Chilton & V. Evans (Eds.) *Language, cognition, and space: The state of the art and new directions*, 51-78. Equinox.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. The MIT Press.

Levinson, S. (2003). Language and mind: Let's get the issues straight. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought*, 25-46. Bradford Book.

Li, P., & Abarbanell, L. (2018). Competing perspectives on frames of reference in language and thought. *Cognition*,170, 9-24 DOI: 10.1016/j.cognition.2017.09.005

Li, P., Dunham, Y., & Carey, S. (2009). Of substance: The nature of language effects on entity construal. *Cognitive Psychology, 58*, 487-524. DOI: 10.1016/j.cogpsych.2008.12.001

Lucy, J. (1992). *Language Diversity and Thought*. Cambridge University Press.

Lucy, J., & Gaskins, S. (2001). Grammatical Categories and the Development of Classification Preferences: A Comparative Approach. In S. Levinson and M. Bowerman (Eds.) *Language Acquisition and Conceptual Development*, 257-83. Cambridge University Press.

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Frontiers in Psychology*, *3*, 54. DOI: 10.3389/fpsyg.2012.00054

Malt, B., Gennari, S., Imai, M., Ameel, E., Tsuda, N., & Majid, A. (2008). Talking about walking: Biomechanics and the language of locomotion, *Psychological Science*, *19*, 232-240. DOI: 10.1111/j.1467-9280.2008.02074.x

Malt, B., Ameel, E., Gennari, S., Imai, M., Saji, N., & Majid, A. (2011). Do words reveal concepts? In *Proceedings of the 33rd Annual Conference of the Cognitive*

Science Society, edited by L. Carlson, C. Hölscher, and T. Shipley, 519-24. Cognitive Science Society.

Majid, A., Boster, J.S., & Bowerman, M. (2008). The cross-linguistic categorization of everyday events: A study of cutting and breaking. *Cognition*, *109*, 235-250. DOI: 10.1016/j.cognition.2008.08.009.

Malt, B., Gennari, S.P., Imai, M., Ameel, E., & Saji, N. (2015). Where are the concepts? What words can and can't reveal. In E. Margolis, S. Laurence (Eds.), *The Conceptual Mind: New Directions in the Study of Concepts*, MIT Press.

Martínez-Manrique, F & Vicente, A. (2015) The activity view of inner speech, Frontiers in Psychology. 6:232. doi: 10.3389/fpsyg.2015.00232

Matchin, W., & Hickok, G. (2020). The cortical organization of syntax. *Cerebral Cortex*, *30*, 1481-1498. DOI: 10.1093/cercor/bhz180

Papafragou, A. (2015). The representation of events in language and cognition. In Margolis, E., & Laurence, S. (Eds.) *The conceptual mind: New directions in the study of concepts*, 327-345. MIT Press.

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition, 108*, 155-84. DOI: 10.1016/j.cognition.2008.02.007

Papafragou, A., & Grigoroglou, M. (2019). The role of conceptualization in language production: Evidence from event encoding. *Language, Cognition and Neuroscience*, *34*, 1117-1128. DOI: 10.1080/23273798.2019.1589540.

Regier, T. & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences, 13*, 439-46. DOI: 10.1016/j.tics.2009.07.001

Regier, T., Kay, P., Gilbert, A. L., & Ivry, R. B. (2010). Language and thought: Which side are you on, anyway? In Malt, B. & Wolff, P. (Eds.) *Words and the Mind: How words capture human experience*, 165-82. Oxford University Press.

Roberson, D., Pak, H., & Hanley, J.R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: evidence from Korean. *Cognition*, *107*, 752-762. DOI: 10.1016/j.cognition.2007.09.001

Slobin, D. (1987). Thinking for Speaking. *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, pp. 435-445. DOI: https://doi.org/10.3765/bls.v13i0.1826

Slobin, D. (1996). From "thought and language" to "thinking for speaking." In . Gumperz, J., & Levinson, S. (Eds.) *Rethinking Linguistic Relativity*. 70-96. Cambridge University Press.

Talmy, L. (1985). Lexicalization patterns: semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description, second edition. Vol. III: Grammatical categories and the lexicon*, 57-149. Cambridge University Press.

Tan, L-H., Chan, A., Kay P., Khong, P-L, Yip, L., & Kang-Kwong Luke, K-K (2008). Language affects patterns of brain activation associated with perceptual decision. *Proceedings of the National Academy of Sciences*, *105*, 4004-4009. DOI: 10.1073/pnas.0708670105.

Thierry, G., Athanasopoulos P., Wiggett A., Dering B., & Kuipers J.R. (2009). Unconscious effects of language-specific terminology on color perception *Proceedings of the National Academy of Sciences*, *106*, 4567-4570. DOI: 10.1073/pnas.0811155106.

Vicente, A., & Jorba, M. (2019). The Linguistic Determination of conscious thought contents. *Noûs*, 53(3), 737–759. https://doi.org/10.1111/nous.12239

Vicente, A & Martínez-Manrique, F. (2016) The nature of unsymbolized thinking. *Philosophical Explorations*, 19:2, 173-187. DOI: 10.1080/13869795.2016.1176234

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, eds. & trans.). Cambridge, MA: Harvard University Press. (Original work published 1930–1935).

Vygotsky, L. S. (1987). Thinking and speech. In *The Collected Works of L. S. Vygotsky*, Vol. 1. New York: Plenum. (Original work published 1934.)

Whorf, B.L. (1956) *Language, Thought and Reality*. Selected Writings of Benjamin Lee Whorf, MIT Press.

Biographical Note

Agustin Vicente received his PhD in Philosophy from the University of the Basque Country. His main field of work has been the relations between language and thought. The funded research projects he has had, have dealt with this issue, and they cover most of the 40 articles published in the last decade (a large part in prestigious international journals such as *Linguistics and Philosophy*, *British Journal for the Philosophy of Science*, *Mind & Language*, *Philosophical Studies* or *Nous*). For this work he had the collaboration of Fernando Martínez Manrique (Granada), Marta Jorba (Pompeu Fabra) and Ingrid Lossius Falkum (Oslo) for quite some time. Currently he also works on autism and language together with Elena Castroviejo Miró (UPV / EHU), with whom he has published also on pejorative language. He published, together with Peter Langland-Hassan, from University of Cincinnati, a book on inner speech at Oxford University Press, *Inner Speech: New Voices*.