CHAPTER 6

Language-Thought Processes

> The limits of my language mean the limits of my world.
> – Wittgenstein, *Tractatus Logico-Philosophicus*

> When a language dies, so much more than words are lost. Language is the dwelling place of ideas that do not exist anywhere else. It is a prism through which to see the world.
> – Robin Wall Kimmerer, *Braiding Sweetgrass: Indigenous Wisdom, Scientific Knowledge, and the Teachings of Plants*

6.1 Introduction

This chapter will look at a type of cognitive process that has been studied by linguists and other cognitive scientists at least since the early twentieth century. Though this type of cognitive process is not explicitly regarded as a cognitive kind, it can be considered to be a candidate for being such a kind, and can be evaluated as such. The type of process I have in mind is one that involves a particularly strong influence of language on thought, or a deep and far-reaching relationship between language and thought. Whether or not there are such cognitive processes is itself a matter of debate, and their very existence has been hotly contested. The issue tends to be framed in terms of subscription to a hypothesis, the "Sapir-Whorf hypothesis," rather than the existence of a particular kind, but in the context of this book, I will consider it to be a question regarding the existence of a particular *kind* of cognitive process, as I will try to explain in the course of this chapter.

The "Sapir-Whorf hypothesis," or sometimes just plain "Whorfianism," is one of the most widely discussed general claims about language. It has attracted some limited attention among philosophers, mainly because of its connection to broader philosophical concerns having to do with conceptual change and conceptual incommensurability (Davidson 1974; Carruthers 2002). Meanwhile, although psychologists and linguists have

investigated the thesis extensively in the past few decades, there does not seem to be a consensus on its actual content or how to formulate it. In this chapter, my aim is to formulate the thesis with greater precision than it appears to have been stated thus far. This will enable us to determine whether it demarcates a real kind of phenomenon or phenomena, and whether the phenomena in question are categorically distinct from others that involve the interaction of language and thought. That, in turn, should help us assess its status as a real cognitive kind. So as not to get embroiled in an exegetical dispute about the actual views of its most famous proponents, Edward Sapir and Benjamin Lee Whorf, I propose to employ a more neutral handle for the thesis that has often been associated with their names, the "Language-Thought (LT) hypothesis." The corresponding candidate for a cognitive kind can be labelled a "LT process" and the psychological effects that are thought to be accounted for by this process can be called "LT effects."

In its simplest form, the LT hypothesis is often stated as the claim that language influences thought or cognition. The thesis is often given in a weak and strong version; sometimes the weak version is labeled "linguistic relativity," while the strong version is termed "linguistic determinism." The weak version is often taken to be the bare claim that language influences thought (to some degree),[1] while the strong version is widely held to be the view that language determines thought (cf. Kay and Kempton 1984; Clark 1996). There is a problem with both versions of this formulation of the LT hypothesis. The former claim is innocuous or indeed trivial: How could there not be some influence of language on thought? (Compare: the view that perception influences thought, or the thesis that emotion influences cognition.) By contrast, the latter claim seems ridiculously inflated: How could language alone determine a person's or a community's entire body of thought or cognitive processes? But if neither of these statements of the position are tenable, is there an intermediate view that is both interesting and plausible, that is neither trivially true nor patently false?[2] The problem has been made more acute by a challenge laid down by two cognitive scientists who have dramatically highlighted the threat of trivialization when it comes to the LT hypothesis. Bloom and Keil (2001) argue that the issue

[1] "Linguistic relativity" is a misleading label for this position, since it suggests merely that languages differ in the way that they describe or categorize features of reality, without making the additional claim that these differences have an impact on thought.

[2] The lack of consensus on the precise content of the thesis is confirmed by the fact that researchers sometimes reach opposite conclusions regarding the support for the thesis based on much of the same evidence (compare e.g. Bloom and Keil (2001) with Reines and Prinz (2009)).

in question cannot be a matter of whether language can have an influence on thought, since: "Nobody doubts that language can inform, convince, persuade, soothe, dismay, encourage, and so on" (Bloom and Keil 2001, 354). Otherwise, they ask rhetorically, "why would you be reading this?" It is an obvious truth that language influences thought every time one cognitive agent uses language to communicate thoughts to another, and this fact threatens to trivialize the LT hypothesis unless we can distinguish this ubiquitous feature of human cognition from the phenomena that are supposed to support the hypothesis.

If one does not take one's cue from the actual views of Sapir and Whorf, how should one proceed to pin down the content of the LT hypothesis with sufficient rigor as to avoid the threat of trivialization and determine whether it has correctly managed to identify a real cognitive kind? I propose to be guided by two sources of evidence, recent empirical work claiming to test the hypothesis, which contains an implicit understanding of its content, and current attempts to state the thesis in a rigorous way. By building on ideas and insights gleaned from these two sources in turn, I aim to formulate it more explicitly and precisely than it has usually been stated. It might seem wrong-headed to survey the empirical evidence before one has looked at attempts to formulate the hypothesis. How can we tell whether the evidence is indeed evidence for the LT hypothesis if we do not yet have a formulation of that hypothesis? The answer is that the empirical cases that I will examine are widely cited as archetypal pieces of evidence for the hypothesis and contain an implicit, if not explicit, understanding of that hypothesis. If it turns out, upon reflection, that some of the evidence cited does not conform to the formulations that we will later encounter, then that evidence can be reassessed, or else the formulations themselves can be revised. The point of the exercise is not merely descriptive, since I will be arguing that some ways of stating the thesis and some implicit understandings of the processes involved are more precise, cogent, and empirically corroborated than others.

The chapter will proceed as follows. In Section 6.2, I will survey some of the empirical evidence that is widely cited to support the LT hypothesis. Then, in Section 6.3, I will examine a number of attempts to formulate the hypothesis, comparing them, pointing out various shortcomings, and highlighting the most promising aspects of each. In Section 6.4, guided both by the empirical results in Section 6.2 and the formulations surveyed in Section 6.3, I will attempt to frame the LT hypothesis in some detail and with sufficient rigor, in such a way that it evades the charge of trivialization and can be used to describe a candidate for a *kind of cognitive*

process. Though it will emerge that there is significant support for the LT hypothesis in recent empirical work, I will argue that the processes demarcated by the LT hypothesis may not be entirely distinct from other types of cognitive process. I will also make the case that they may not constitute a homogeneous collection of cognitive processes in their own right. The upshot will be that Language-Thought Processes do not constitute a cognitive kind, but rather that they are of two distinct types of process, each of which may be subsumed within a broader cognitive kind.

6.2 Empirical Evidence for the LT Hypothesis

Writing a few decades ago, Kay and Kempton (1984, 67) asserted that "the bulk of the [empirical] research" designed to test the LT hypothesis "has concerned the domain of color." Fortunately, this assessment has been made obsolete in the interim, since the past few decades have witnessed a flurry of empirical activity that has examined the LT hypothesis in a number of different domains. Consider a few paradigmatic experimental results that are meant to support the LT hypothesis, drawn from different areas of cognition:

a) ***Color:*** Russian speakers, who use different terms for light blue (*goluboy*) and dark blue (*siniy*), can more easily discriminate two shades taken from the two distinct categories than two shades taken from the same linguistic category; English speakers do not show this effect. Thus, linguistic categories affect performance on a nonverbal perceptual task. The effect for Russian speakers can be disrupted by a verbal interference task, apparently confirming the influence of language on thought (Winawer, Witthoft, Frank, et al. 2007).[3]

b) ***Spatial coordinates:*** Speakers of languages that primarily employ absolute spatial coordinates (e.g. Guugu Yimithirr, Tzeltal) carry out nonlinguistic spatial tasks differently than speakers of languages that primarily employ relative spatial coordinates (e.g. English, Dutch). Specifically, the former tend to duplicate visual scenes and remember them in a way that accords with cardinal directions, while the latter do so in accord with egocentric directions (Levinson 2003; Majid, Bowerman, Kita, et al. 2004).

[3] Since color is perhaps the most active research focus in this broad area of inquiry, there is obviously a great deal more to be said about the influence of color language on color perception. See Cibelli, Xu, Austerweil, et al. (2016) for a model of color cognition that combines a universal color space with language-specific partitions of that space.

c) ***Gender:*** Spanish and French speakers show effects of grammatical gender on classification when asked to assign either a woman's voice or man's voice to an object in a picture (supposedly for an animated movie). These effects emerge reliably only around 7 years of age, indicating that there is a "time lag between language acquisition and the infiltration of language into the cognitive system …" of the order of several years (Sera, Elieff, Forbes, et al. 2002; cf. Boroditsky, Schmidt, & Phillips 2003).

d) ***Categorization:*** English-speaking and Japanese-speaking children and adults generalize object instances and substance instances differently, suggesting an influence of language on categorization, having to do with differences in the use of count-mass terms and classifiers in English and Japanese. The differences are more pronounced among adults than among children (Imai & Gentner 1997).

e) ***Motion:*** Direction of motion is more often conveyed by the verb itself in Spanish and by a preposition in English, whereas manner of motion is usually encoded in the verb itself in English and in other supplemental expressions (e.g. adverbs) in Spanish. Correspondingly, children who speak English describe an illustration from a storybook differently from their Spanish-speaking counterparts, and the effect is more pronounced at age nine than at age five. It has been suggested not only that native English and Spanish speakers describe a visual scene differently due to linguistic differences between the two languages, but that the process of "thinking for speaking" eventually leads children to notice different things about a visual scene (Slobin 1996; cf. Slobin 2003).

Although all these results have been widely cited, none of them are without their critics and detractors. But the criticisms tend to claim that the evidence does not adequately support the conclusion, not that the conclusion is incoherent or not well formulated. For example, the study cited in example (e), above, has been criticized on the grounds that it does not show that children notice different things in a picture but that they simply describe what they notice differently (Gennari, Sloman, Malt, et al. 2002, 55). However, where these results have been accepted, they have been taken to demonstrate LT effects. Hence, it seems warranted to regard them at least as a tentative guide to what genuine LT effects would amount to, and hence what a LT process would be.[4]

[4] There is another class of effects and body of research that pertain to a generalized version of the hypothesis that language influences thought. This version states that having language endows humans with

https://doi.org/10.1017/9781009223645.007 Published online by Cambridge University Press

6.3 Formulations of the LT Hypothesis

Now that we have sampled some representative empirical results that have been judged to support the LT hypothesis, we should have a better idea of what we are looking for. The common denominator among these results would seem to be that they all show evidence that the native language[5] of a speaker has some influence on other, nonlinguistic aspects of the speaker's cognitive capacities (e.g. perception, memory, spatial cognition, categorization). The influence is, presumably, causal, and the effects are supposed to obtain even when these aspects of a speaker's cognitive capacities are tested on tasks that do not require explicit linguistic representation. To be sure, in some of the experiments mentioned in the previous section, language is involved in some way in performing the task, but the aspects of language that are under examination are assumed not to be explicitly recruited by the task demands.

There are few attempts in the literature to formulate the LT hypothesis in any detail; many research articles on the topic content themselves with a paraphrase to the effect that language influences thought or nonlinguistic cognition in some way. Indeed, some formulations of the LT hypothesis resort to figures of speech, employing the metaphors of language "shaping" or "molding" thought, without attempting to describe the type of process in any detail. The four formulations to be discussed here constitute exceptions in that they attempt to articulate the LT hypothesis in more detail and in a nontrivial fashion.

certain cognitive capacities that they would have otherwise lacked. This general linguistic capacity (as opposed to specific natural languages) is supposed to be at least partly responsible for some of the cognitive abilities of human beings that set them apart from other animals (see e.g. Dennett 1997; Clark 1998; Gentner 2003; Spelke 2003). This thesis seems weaker than the LT hypothesis, in the sense that one could hold it while not accepting LT, though the converse may not be true. If one holds that having language, as opposed to having no language at all, has an influence on thought, it does not follow that mastering, for example, Arabic as opposed to English has an influence on one's thought. By contrast, if it is true that being a native or competent speaker of some specific language has such an influence, then it would follow that having language generally influences thought. Since it is a weaker thesis, I will not discuss the general thesis further here, though formulating the specific LT hypothesis may also help to articulate the precise content of the more general thesis. In Chapter 2, I discussed the influence of language on concept possession and acquisition, arguing that linguistic symbols are instrumental in determining conceptual identity, and this point will be taken up again in Section 6.4.

[5] Why does it have to be one's native language? Presumably because effects associated with one's native language are more far-reaching than those associated with other languages that one might have mastered. Hence, if there is an effect at all, it ought to be more noticeable in the case of a thinker's native language. But see, for example, Dolscheid, Shayan, Majid, et al. (2013) for LT effects resulting from training native Dutch speakers on a linguistic metaphor in Farsi. But they posit that "participants received a concentrated 'dose' of the relevant linguistic metaphor, probably equivalent to weeks or months of normal language use" (Dolscheid, Shayan, Majid, et al. 2013, 619). Thus, these effects may pertain more properly to proficient or competent speakers rather than native speakers.

6.3.1 Bloom and Keil

In an attempt to respond to their own challenge regarding the trivialization of the LT hypothesis, Bloom and Keil (2001, 354) state that the real question is not whether language can have an influence on thought, since this occurs every time we use language to communicate with one another. Rather, the issue is "whether language shapes thought in some way other than through the semantic information that it conveys," that is, "whether the structure of language – syntactic, morphological, lexical, phonological, etc. – has an effect on thought." But this response does not seem quite right for two reasons. First, LT effects do not always involve structural differences among languages, even if we understand linguistic structure more broadly than just in terms of syntax. In the case, for example, of the experiments concerning Russian speakers' perceptual discrimination of shades of blue, the influence does not have to do with the structure of the Russian language, merely with having two words to denote different shades of blue (example (a) in Section 6.2). At the very least, it is debatable whether we should consider that Russian's having an additional lexical item relative to English is a difference in lexical structure. The second problem with Bloom and Keil's proposal is that these linguistic differences often do involve semantic information. A Russian speaker who has an additional lexical item for a shade of blue can be considered to have additional *semantic* information, for example, she knows that *siniy* refers to *that* shade. Similar remarks apply to languages with absolute as opposed to relative coordinates (example (b) in Section 6.2), which supply language users with additional conceptual resources, without affecting the very structure of the language.[6] But there may be a way of modifying Bloom and Keil's proposal to make it fit the phenomena. In all the cases sampled in the previous section, the speaker is equipped with one or more additional lexical or morphological items to stand in for certain concepts or conceptual constituents. Rather than say that the structure of language influences thought in these cases, it may be more accurate to say that certain aspects of language (primarily lexical or morphological aspects) influence certain habits of thought or cognitive tendencies. In fact, the issue does not appear to be about whether the *structure of language* influences thought, but rather whether language influences the *structure of thought*, such as categorization abilities or perceptual discrimination. But

[6] Even examples (c), (d), and (e) in Section 6.2 are not purely syntactic in nature, but combine syntax and semantics. It is true that in some of Sapir and Whorf's earlier work, syntactic differences were emphasized over semantic ones, but the bulk of the empirical work since then suggests that both kinds of LT effects are attested, and indeed are hard to disentangle.

this constitutes a rather vague response to Bloom and Keil's challenge, so I will try to make these locutions more precise in what follows.

6.3.2 Carruthers

Another attempt to delineate the phenomena of interest comes from Carruthers (2012), who observes that traditional "Whorfianism" held that language has a "*structuring* effect on cognition," thereby seeming to confirm the variation on Bloom and Keil's formulation proposed in Section 6.3.1. But he goes on to explicate this idea by saying that it holds that "the absence of language makes certain sorts of thoughts, or certain sorts of cognitive process, completely unavailable to people" (Carruthers 2012, 385). This claim, which Carruthers rejects, accords with the strong version of the LT hypothesis mentioned earlier, which states that language determines thought in the sense that it prohibits (or mandates) the thoughts that are available to speakers. After rejecting this discredited view, he goes on to argue that a more plausible view has recently been gaining ground, namely that "natural language can make certain sorts of thought and cognitive process more *likely*, and more *accessible* to people" (Carruthers 2012, 385; original emphasis). The idea that language makes certain thoughts or cognitive processes more likely and accessible appears to accord with the experimental results vetted in Section 6.2. Indeed, Carruthers to the contrary notwithstanding, this would seem to be one way of elaborating on the vague notion that language "structures" cognition. One obvious way to test for whether cognition has been restructured is to gauge whether certain cognitive processes have been rendered more likely to occur and more accessible to language users. If our cognitive abilities have indeed been restructured, then this should show up in the cognitive processes that we engage in, especially in making certain responses more or less likely, or in making certain thoughts more or less accessible. Hence, we might do well to retain the idea of language restructuring thought, at least provisionally, in attempting to formulate the LT hypothesis, when understood in terms of making some thoughts more likely or accessible than others. This is also consistent with the claim that the influence of language on thought does not involve either mandating or prohibiting certain thoughts or habits of thought outright. In other words, it agrees with the widespread consensus that to the extent that the LT hypothesis is defensible, it is so in its weak rather than its strong form. In contrast with strong LT, weak LT talks about facilitating (or inhibiting) rather than mandating (or prohibiting) cognitive processes, in the sense of making them more or less likely or more or less accessible.

6.3.3 Hunt and Agnoli

In a somewhat earlier discussion, which remains relevant, Hunt and Agnoli (1991) put forward some concrete suggestions for thinking about the ways in which language might restructure thought. "The weaker form of the [LT] hypothesis states that language differentially favors some thought processes over others, to the point that a thought that is easily expressed in one language might virtually never be developed by speakers of another language" (Hunt & Agnoli 1991, 378). This agrees with the idea encountered above (in Section 6.3.2) that some thoughts may be rendered more probable and accessible for speakers of some languages as opposed to others. Hunt and Agnoli expand on this by saying that some thoughts will be more "natural" and "come easily" for the speakers of some languages relative to others, though they admit that these notions are not capable of a scientific construal. However, they go on to make some interesting suggestions as to how one might render the contrast more amenable to empirical investigation. Using the example of southern Californian surfers who have recruited words like "hollow" and "flat" to describe specific types of ocean waves, Hunt and Agnoli (1991, 378) claim that a language user who lexicalizes a concept has "traded expensive space in short-term memory for cheaper space in long-term memory." Presumably, even though non-surfers may be able to recognize the relevant differences among waves and recall them on a later occasion, surfers may be more efficient at doing so, partly because they have convenient labels that can be recruited, whether overtly or covertly, to do so. Lexicalizing the concept in a language may make it more probable that the concept is accessed by a thinker, or make it accessible more efficiently. Moreover, it may make it possible to access it while performing other, nonlinguistic cognitive tasks that make demands on short-term memory. This, then, may be one process whereby cognition may be restructured by language, or some thoughts rendered more probable and accessible as a result of the introduction of linguistic expressions for them.

But though this is an illuminating discussion, it does not yet answer the Bloom-Keil challenge, since the question still remains: How do these kinds of facilitatory effects differ from the ways in which language routinely promotes or encourages thoughts in ordinary human communication? It may be useful to compare two cases of human communication to better understand the difference. Consider first a case in which I ask a first-grader to turn the lights off when she leaves the bathroom after brushing her teeth. She assents, and minutes later, she complies with

my request. Here, a linguistic utterance has influenced the thoughts of my interlocutor, leading to her subsequent action. Now imagine a case in which I explain to the same first-grader that the planet's climate is changing as a result of human activity, perhaps introducing her to such concepts as *climate* and *global warming*. Then that may encourage her to think certain thoughts, to the effect that she should take steps to reduce her own energy use and that of others. It may make it more likely for her to think of switching off the lights when she leaves a room, not just on a specific occasion but on many future occasions. It may even entail a significant restructuring of her thoughts about her relationship to her environment, her consumption habits, diet, lifestyle, and so on, prioritizing thoughts about energy conservation as she makes daily choices about transportation, eating, and entertainment. In both cases, I have used language to influence the thoughts (and actions) of the child. But it is clear that in the second case, the influences are more far-reaching and enduring. There may be no strict dividing line between the first type of case and the second, and there may be intermediate cases (for example, a case in which I explain to the child that it is a waste of energy to keep lights on in an empty room without going into the ramifications for the planet's climate). But it seems clear that there is a significant difference between the first type of case and the second, and that part of the difference has to do with lexicalizing concepts like CLIMATE and GLOBAL WARMING. Once these concepts are lexicalized, they are stored in long-term memory, as Hunt and Agnoli observe, and may have a more significant impact on the cognitive life of the agent in the long run, making some thoughts generally more likely and more accessible.

It may be objected that in the case of explaining climate change, the child is not learning a natural language but is instead learning a scientific theory, and so the effect on cognition results from acquiring a specific body of knowledge, not simply from being proficient in a natural language. Hence, we might stipulate that LT effects are those that affect a language user merely in virtue of being a competent speaker of that language, not in virtue of learning a new theory or a new set of concepts in that language. Part of what distinguishes LT effects from standard communicative acts may be said to be that they involve mere competence in a cognitive agent's natural language, rather than familiarity with a specialized domain of knowledge. In genuine LT effects, it appears that it is important that the effect in question result merely from the fact that the language speaker is a competent speaker of his or her native language. (However, this assumption will be reexamined in Section 6.4.)

6.3.4 Wolff and Holmes

Finally, a more recent formulation of the LT hypothesis comes from Wolff and Holmes (2011, 261), who use a metaphor to convey the gist of the thesis: "There is evidence … that while language may not close doors, it may fling others wide open. For example, language makes certain distinctions difficult to avoid when it meddles in the process of color discrimination or renders one way of construing space more natural than another." This suggests that the LT hypothesis construes language as a promoter (or inhibitor) rather than a dictator (or prohibitor), making some thoughts more or less likely (rather than mandatory or unattainable), as we have already concluded. But (again) this is not sufficient to answer the challenge conjured up by Bloom and Keil. Wolff and Holmes (2011) go on to provide a useful taxonomy of the ways in which language can "open doors." The main categories in their taxonomy are as follows: (i) thinking before language, (ii) thinking with language, and (iii) thinking after language, and they provide examples of each of these categories. In the first category, they place the phenomenon of "thinking for speaking," according to which, for example, English- and Spanish-speaking children notice different things about a picture because of the need to put it into words in their respective languages (Slobin (1996); example (e) in Section 6.2). In the second category are such phenomena as the color discrimination tasks with Russian and English speakers, in which Russian speakers are better at discriminating color shades that are labeled differently in Russian (Winawer, Witthoft, Frank, et al. 2007; example (a) in Section 6.2). Meanwhile, the third category includes cases involving differences between speakers of languages with absolute coordinates and speakers of languages with relative coordinates, specifically differences in how they replicate a visual scene (Majid, Bowerman, Kita, et al. 2004; example (b) in Section 6.2). This classification scheme is based mainly on the temporal sequence in which language and thought interact, whether the influence of language on thought occurs before, during, or after the cognitive process. But that cannot be quite right, since it is incoherent to say that language has an influence on the cognitive process *after* that process has occurred. Rather, if one lumps together (i) and (ii), and contrasts them with (iii), the distinction seems to be between two kinds of influence, one that is more time-sensitive, in which language is covertly implicated in a particular thought process as it occurs or shortly before it occurs, whereas the other has to do with changes in cognition that have been effected over time as a result of habitual language use and persist even in the absence

of language use. In the first two broad categories, the interaction between thought and language is simultaneous or nearly so. Even when performing some nonverbal tasks, there is evidence that thinkers use language as a tool or crutch and that it induces them to think certain thoughts or engage in certain cognitive processes. The paradigmatic examples of these cases may be those that involve lexicalizing a concept, allowing us to store it in long-term memory and access it more reliably in carrying out certain cognitive tasks, such as perceptual discrimination tasks. In the third category are phenomena that implicate ingrained habits of thought that are the result of linguistic influences that are more long-term than the other two. These cases would seem to be different in that they typically involve associations among concepts, which are made as a result of habitually using a language and of making certain distinctions required by some natural languages (though not others). Using a language that makes gender distinctions may encourage speakers of that language to make associations between gendered nouns for inanimate objects and stereotypical properties associated with males and females in that society (as in example (c) in Section 6.2). Therefore, there would appear to be at least two kinds of cognitive process involved in typical LT effects. The first, which corresponds roughly to (i) and (ii) in the above taxonomy, involves long-term memory storage of certain distinctions that are lexicalized in one's native language. The second, corresponding to (iii), involves making habitual associations between aspects of reality that are associated by one's native language. As Wolff and Holmes observe, some LT effects are typically erased or diminished by verbal interference but others are not. The reason is that in the type-(i) and type-(ii) cases, language is thought to be covertly recruited in performing the task, whereas in the type-(iii) cases, language need not be recruited simultaneously, since it is posited to have had the effect on cognition after a long period of use, and the effect is not cancelled by temporarily disrupting our linguistic cognitive processes through verbal interference tasks. These latter cases may include those experimental results surveyed in Section 6.2 in which the effect appeared or became more pronounced only several years after learning language, or even in adulthood (examples (c), (d), and (e)).

A great deal more work would have to be done to understand the processes that underlie these effects, but at a first pass, there are likely to be two distinct kinds of process responsible for these two types of LT effects. In the first kind of process, long-term storage is involved, whereby linguistic labels enable us to store items in long-term memory rather than having to keep them in short-term memory, thereby saving precious cognitive resources. This would explain why some items can be recalled

more efficiently and reliably. It may also explain why linguistic interference would disrupt this task, since one's language faculty may need to be recruited to facilitate performing the recall task, which is mediated by the use of a linguistic label. As for the second kind of process, it is likely to be of the associative sort, which appears elsewhere in cognition. In this case, frequent association between concepts, due to language use, leads to a strong connection between these concepts, even when the concepts are not recruited by a task involving language use. This distinction will be explored further in the next two sections.

6.4 Proposal and Discussion

Now that we have surveyed several prominent recent attempts to pin down the LT hypothesis, it is possible to derive the most promising elements from these attempts in order to try to emerge with a more precise formulation. The LT hypothesis can be said to consist of the conjunction of the following four claims concerning particular natural languages:

1) Language *facilitates or inhibits* (rather than dictates or prohibits) a speaker's thoughts or cognitive processes;
2) Language effects *structural cognitive modifications*, making certain thoughts more (or less) accessible, in cognitive tasks involving perceptual discrimination, memory, and categorization, among other cognitive processes;
3) Language is implicated in the effect merely as a result of the agent being a *competent speaker* of that language, not as a result of acquiring additional information, such as expert knowledge;
4) Language (i) has a *near simultaneous* and covert causal influence on cognition as a result of lexicalizing one or more concepts, or (ii) produces a *nonsimultaneous and long-term* causal change in cognition as a result of habitual associations.

Even if they do not constitute necessary and sufficient conditions, these can be considered the main tenets of the LT hypothesis, and the paradigmatic experimental results sampled in Section 6.2 seem to bear them out. But do they enable us to answer the Bloom-Keil challenge, and do they delineate a homogeneous and distinct class of cognitive effects? When it comes to the former question, some of these clauses, particularly (3), rule out the kind of communicative processes by means of which language routinely affects thought as in standard communicative acts. But it may be objected that we have, in effect, *defined* the phenomenon in such a way

https://doi.org/10.1017/9781009223645.007 Published online by Cambridge University Press

that the Bloom-Keil challenge does not arise. Perhaps, but the initial task was to delineate this class of effects, not to justify why these effects ought to be distinguished from other cognitive phenomena involving language. As I will try to argue presently, there is unlikely to be a principled difference between those cognitive effects that result from being a competent speaker and those that ensue from acquiring expert knowledge. This formulation of the LT hypothesis may accord with the paradigm cases and may indeed capture those phenomena that have commonly been considered LT effects, but that does not mean that LT effects constitute either a homogeneous or a distinct category of cognitive phenomena. Now that we have a delineation of the phenomena, we can go on to consider whether there are principled reasons for focusing on this class of effects. Do they constitute a homogeneous and distinct category of effects, in other words, are they the effects of a *kind* of cognitive process? In the rest of this section, I will try to argue that not only are LT effects as a whole not likely to be distinct from a broader set of cognitive effects, they are also likely to issue from two different kinds of cognitive process.

It will be convenient to proceed by considering each of the claims (1)–(4), above, starting with (1), the idea that LT effects are facilitatory or inhibitory, rather than mandatory or prohibitory. This aspect of the LT hypothesis reflects the consensus that the version of the hypothesis that has strongest empirical support is the weaker rather than the stronger version, sometimes known as "linguistic relativity" (as opposed to "linguistic determinism"). The existing evidence suggests that whatever cognitive effects language may have, they neither mandate nor prohibit certain thoughts or cognitive operations. Some philosophers and cognitive scientists do claim that some languages are incommensurable with others, but they would still allow that thinkers can make the transition from one language to the other or learn the second language (Carey 2009). Presumably, these language learners can acquire concepts that are not lexicalized in their native language, even though they may have to overcome certain cognitive hurdles to do so. Hence, (1) is a fairly well-established component of the LT hypothesis and seems to be supported by the effects we have surveyed.

As for (2), the claim that language effects certain structural modifications in one's cognitive system, this has been explicated in terms of making certain thoughts more likely or accessible (though that may be understood as a way of testing or operationalizing the claim). One process for achieving a restructuring is the one proposed by Hunt and Agnoli (1991), who contend that lexicalizing a concept may involve (among other things) storing it in

long-term rather than short-term memory.[7] This would facilitate access to it and ensure that it can be used in a cognitive task that places demands on short-term memory, as in tasks involving remembering the spatial configuration of objects (example (b) in Section 6.2). But there may also be other cognitive processes that yield LT effects, such as habitual associations built up between concepts as a result of the fact that the corresponding words or morphemes are associated in natural language. For instance, French speakers may associate the word for an inanimate object that takes the masculine gender with stereotypical masculine traits and hence be more prone to associate that object with a male voice rather than a female voice (example (c) in Section 6.2). Speakers of languages in which the word for the same object takes the feminine gender or is gender-neutral may not exhibit the same tendency because they have not formed the habitual association between the object in question and the stereotypical gender traits. The processes that subserve these associative effects seem different from those that would be involved in storing a lexical item in long-term memory. They need not be disrupted by verbal interference and may take longer to become entrenched. Despite the different processes involved, both types of LT effects can be thought of as restructuring cognition in the sense of making certain cognitive responses more accessible or more likely.

When it comes to clause (3), this may be the most controversial aspect of the above characterization of the LT hypothesis. This clause was used to distinguish standard LT effects from instances of concept acquisition leading to cognitive restructuring, as in the case of a child whose habits of thought change as a result of learning about climate change, acquiring such concepts as CLIMATE and GLOBAL WARMING. We might distinguish these phenomena from LT effects by saying that they do not involve the acquisition of a natural language, but rather the acquisition of a novel set of concepts within a language. But the problem with this suggestion is that the line between languages (or language fragments) and conceptual repertoires may not be a principled one. Indeed, some of the examples that have been used to illustrate the phenomenon of cognitive restructuring as a result of LT effects involve the acquisition of concepts by speakers within the same natural language, as in the example of California surfers lexicalizing certain wave formations (mentioned in Section 6.3.3), rather than a contrast between speakers of different natural languages. It is well

[7] Presumably, it also somehow involves establishing a link between long-term semantic memory and a particular lexical item or morpheme, since lexicalization is not a mere matter of storage in semantic memory.

established that within the same natural language, experts tend to deploy concepts and categories that novices do not, for example, for describing different tastes of wine (Solomon 1990), or for classifying different types of physics problem (Chi, Feltovitch, & Glaser 1981), among many other features of reality. LT effects do not seem different in principle from expert-novice effects, for these also influence perception, memory, and other nonlinguistic cognitive functions, as is now well established, and as some proponents of the LT hypothesis have themselves observed (cf. Majid 2002). Moreover, they would appear to involve the same process of lexicalization leading to storage in long-term memory. The difference between the deployment of additional concepts by experts (compared to novices) within a single natural language and the deployment of distinct concepts by speakers of different natural languages is not a principled one from the psychological or linguistic point of view. Indeed, if the linguistic differences between experts and novices are pronounced enough, they can be considered to speak different dialects, and it is a platitude that the difference between languages and dialects is hardly principled either. This seems to leave us with no clear way of distinguishing at least some LT effects from some expert-novice effects in which concepts are lexicalized in a specialized vocabulary. What seems more relevant is the extent to which the acquisition of information leads to cognitive restructuring (as in clause (2) above). This kind of restructuring is far more likely to result from acquiring new concepts or a new theory (e.g. about global climate change) rather than the simple exchange of information in an ordinary communicative act. But the restructuring can just as well be the result of acquiring and lexicalizing novel concepts as it can be the result of mere competence in a natural language.

A different sort of problem with this characterization of LT effects has to do with clause (4), which would seem to pick out two distinct classes of phenomena, those in which language is simultaneously and covertly recruited during a cognitive process and those in which language has a long-term effect on cognition due to habitual association. Now there are some researchers who tend to regard the former as not being genuine LT effects, and some experimental results are summarily dismissed or explained away by saying that they involved "inner speech" or some unobserved recruitment of the linguistic capacity during the completion of a cognitive task (see e.g. Bloom & Keil 2001, 356; cf. Casasanto 2008, 70; Casasanto 2016, 159–160). Other researchers hold that explicit recruitment of language during the completion of a nonlinguistic task is at least as significant as habitual associations that are reinforced by an extended period

of language use (see e.g. Lupyan, Abdel Rahman, Boroditsky, et al. 2020). But there is no need to discuss which of these two types of effect is more important, since the crucial point for our purposes is that there appear to be two types of cognitive process in play, corresponding to two different putative cognitive kinds. In the next section, I will try to further articulate the difference between these two cognitive processes, based on current empirical findings and taxonomic practices in the cognitive sciences.

6.5 Two Kinds of LT Process

Over the past decade, empirical work on LT effects in cognitive science seems to be converging on the conclusion that there are two distinct types of cognitive process involved in producing such effects. A number of researchers have tried to articulate the differences between these two kinds of cognitive process, as can be seen from the following remarks by contributors to this research program:

> … language affects performance differently in different cases. In the case of color discrimination [e.g. Winawer, Witthoft, Frank, et al. 2007], language appears to be having an acute effect; it is involved online, meddling in perceptual decisions as they are made. In the case of grammatical gender [e.g. Boroditsky, Schmidt, & Phillips 2003], language appears to be more of a chronic affliction; it has had a long-term effect on the underlying conceptual representations such that even if language is disabled, its effects in the conceptual system remain. (Boroditsky 2012, 625)

> One possibility is that these [cross-linguistic perceptual] differences stem from long-term perceptual learning caused by years of distinguishing colors in one language ... An alternative is that the cross-linguistic perceptual differences arise from online top-down influences of language … (Perry & Lupyan 2013, 122)

> The present results support the proposal that language can also influence people's low-level perceptuomotor abilities … and that cross-linguistic differences in mental representation can be observed even when people are not using language on-line, overtly or covertly. (Dolscheid, Shayan, Majid, et al. 2013, 619)

> We begin by describing two types of linguistic effects on perception: offline effects in which long-term experience with a specific language affects how people subsequently experience certain perceptual inputs, and online effects in which some aspect of language, such as an object's name, interacts with in-the-moment visual processing. (Lupyan, Abdel Rahman, Boroditsky, et al. 2020, 931)

These observations confirm that current empirical evidence points to the conclusion that all language-thought effects should not be lumped

https://doi.org/10.1017/9781009223645.007 Published online by Cambridge University Press

together in the same category, but rather that they divide into (at least) two distinct taxonomic categories (an argument also made by Wolff & Holmes (2011), as seen in Section 6.3.4).

In discussing the various clauses of the formulation of the LT hypothesis in Section 6.4, we have been led to acknowledge that it neither delineates a class that is distinct from other cognitive phenomena (i.e. ones not usually classified as LT effects) nor does it pick out a homogeneous class of cognitive phenomena. It does not pick out a distinct class because the cognitive effects of conceptual disparity across natural languages are not in principle different from those of conceptual innovation within the same natural language. A Californian surfer equipped with additional terms to describe different types of ocean waves is not in principle different from a Russian speaker who has two terms to describe different shades of blue, relative to a speaker of (standard) English. Still, Bloom and Keil's trivialization problem is not vindicated, since the cognitive effects associated with LT phenomena are not akin to simple communicative acts using language, but are perhaps more closely aligned with those of conceptual change or conceptual enrichment, as in the acquisition of new scientific concepts or a novel set of expert categories. We cannot summarily dismiss these phenomena as non-LT effects by adding a clause to the effect that LT effects involve mere competence in a natural language, since there is no clear distinction between linguistic competence in a natural language and mastery of specialized vocabulary in that language. But in addition to the fact that LT effects may not be clearly demarcated from other cognitive phenomena, these effects may not be coherent or homogeneous, since those effects that involve the simultaneous recruitment of language ("online") are likely to result from different cognitive processes than those that do not ("offline"), though neither of these processes is currently understood well enough to enable us to say for sure. The terms "online" and "offline" are convenient labels that have been used in some recent work to denote the difference between these two kinds of causal process (e.g. Casasanto 2016).[8] But though they are suggestive metaphorical terms, they should not hide the fact that the precise causal pathways in question are not yet well understood. Still, it may be worth summarizing a few of the main features of each type of process, as currently conceived.

[8] See also Boroditsky (2012), Lupyan (2012), Perry and Lupyan (2013), Dolscheid, Shayan, Majid, et al. (2013), and Lupyan, Abdel Rahman, Boroditsky, et al. (2020), most of which were quoted above. But note that Slobin (2003) seems to use the term "online" for processes that others would label "offline."

The first type of cognitive process entails lexicalizing a concept and storing a lexical item in long term memory. This item is then thought to serve as a tag or label that is covertly deployed in the course of performing some nonlinguistic task, such as a perceptual or memorial task, facilitating the performance of that task. Lupyan (2012) has proposed the "label-feedback hypothesis" to account for this type of language-thought effect, though he admits that it is just a sketch. As he explains it, visual representation of the color blue, say, can spontaneously activate the word "blue," and this in turn modulates the visual representation. This means that linguistic labels can be covertly deployed in the course of performing some nonlinguistic task, often providing a way of carrying it out more accurately or rapidly, or at any rate differently from the way in which it would be carried out without dependence on the linguistic label. Some researchers think that this may involve "subvocal linguistic encoding" during the performance of the task (Bohnemeyer 2020), while others posit that the "phonological loop," widely considered to be a constituent of short-term or working memory, has a role to play in executing such tasks (Athanasopoulos & Casaponsa 2020). This is why such effects are thought to be extinguished under conditions of verbal interference.

The second type of cognitive process typically operates over a longer time period, since it involves the forging of habitual associations between two or more concepts as a result of prolonged language use. Different languages represent concepts differently. In the simplest case, some languages may lexicalize certain concepts, while others may not. This means that over a long period of language use, the features associated with those concepts will be made more salient for users of some languages rather than others, and they will tend to notice them more frequently, be more efficient at categorizing instances falling under them, and otherwise perform differently on certain cognitive tasks. Moreover, these tasks need not involve language use, whether overt or covert, as long as they result in significant effects on overall cognition. Similarly, some languages make associations between concepts that other languages do not. For example, while Dutch describes acoustic pitches as high or low, Farsi describes them as thin or thick, thus forging associations between these acoustic and spatial concepts. In one study, Dutch and Farsi speakers were found to perform differently on a nonlinguistic task requiring them to reproduce musical pitches, in line with the spatial metaphors employed in their respective languages. The researchers concluded that "speakers of different languages tend to form systematically different representations of the same physical experiences, even when they are not using language" (Dolscheid, Shayan, Majid, et al.

2013, 620). This claim may be questioned on principled grounds, since some researchers have doubted that effects of language use could somehow "reach down" and affect such abilities as perception and memory (e.g. Pinker 1995). But there is ample evidence of top-down effects in other areas of cognition and it is well established that language use can lead to changes in perceptual discrimination abilities. To take just one relevant example from phonology, very young infants initially discriminate consonant contrasts that are not made in their native languages, but by the age of ten to twelve months old, they no longer make these contrasts and perform similarly to the adults in their linguistic communities (Werker & Tees 1984; cf. Gleitman & Papafragou 2005). This is an effect of phonology, not syntax or semantics, on auditory perception, but it demonstrates that even relatively limited exposure to language can have an influence on perceptual discrimination abilities.[9] Other effects of language use on thought can proceed in a similar fashion by affecting preexisting innate predispositions. Language use may not create these dispositions but modify preexisting dispositions, either enhancing or diminishing them.

To summarize, there is considerable evidence to suggest that there are two types of process that result in what are generally taken to be LT effects. Since they operate on different time-scales and exploit different cognitive resources (e.g. covert labelling, habitual associations), it is reasonable to conclude that these processes mark out different causal pathways, and hence, correspond to different cognitive kinds. As mentioned earlier, different experimental paradigms have been developed to detect these processes, and these include verbal interference designs intended to disrupt the simultaneous and covert recruitment of linguistic labels during the performance of a nonlinguistic task, thereby distinguishing online processes from offline ones. These experimental paradigms have been used to support the hypothesis that there are two cognitive kinds at issue. Some researchers suggest that the ultimate vindication of this claim would lie in the discovery of neural mechanisms that are proprietary to each cognitive process. Neural mechanisms may yet be discovered that pertain to each of these processes, and while reduction of each cognitive process to a type of neural process cannot be ruled out, it should not be assumed either. Unlike some other cases analyzed in previous chapters, in the case of LT processes, there does not seem to be an obvious obstacle to the identification

[9] In the study by Dolscheid, Shayan, Majid, et al. (2013) on musical pitch reproduction in Dutch and Farsi speakers, they trained Dutch speakers to employ Farsi-like spatial metaphors for pitch, as a result of which Dutch speakers' performance resembled that of Farsi speakers. See also footnote 5, above.

of a type reduction of cognitive kinds to neural kinds. Even though LT effects concern the conceptual contents associated with linguistic terms, and I argued in Chapter 2 that conceptual content is individuated (in part) externalistically, what appears to be relevant in these cases is the internalist component of the conceptual contents. In other words, these cognitive effects pertain to language users' conceptions, not the public meanings of concepts in their languages or dialects. Hence, the two posited types of LT process in question may have types of neural counterparts.

These effects of language on thought can help illuminate the ways in which language does not merely act as a medium to convey and communicate thought, but enters into the structuring and shaping of thought itself. These structural effects on thought corroborate the claim made in Chapter 2 that language is not just a passive medium for the expression of concepts but rather enters into the very individuation of concepts. Moreover, this understanding of LT effects and processes may shed some light on what can be termed "general" LT effects. A weaker but more general thesis associated with language and thought is that language possession (as opposed to not having language at all) helps equip creatures with certain cognitive abilities that they otherwise would not have had.[10] The previous discussion of two kinds of LT process can provide some further understanding of how this is possible. The first type of "online" LT process is one whereby a language user deploys a label to help them perform a nonlinguistic task, such as a perceptual or memorial one, and this could explain why language users might have an advantage in performing some cognitive tasks that do not involve language. Cognitive agents who use language can avail themselves of this labelling device even when performing nonlinguistic tasks, involving perception, learning, and memory. The second type of "offline" LT process results from associations formed on the basis of extensive language use, and can also assist with certain cognitive tasks by strengthening associations among phenomena that are correlated in reality. If the existence of these processes is corroborated by further inquiry, then they could also explain why language users might have an edge over nonusers when it comes to their ability to engage in solving certain problems in the cognitive domain. Of course, the cognitive impact of being a language user need not be limited to these effects, but they may at least be part of the picture (see Figure 6.1 for a diagrammatic representation of the relationship to language-thought processes to processes of conceptual change and indicating their division into offline and online processes).

[10] See footnote 4 above.

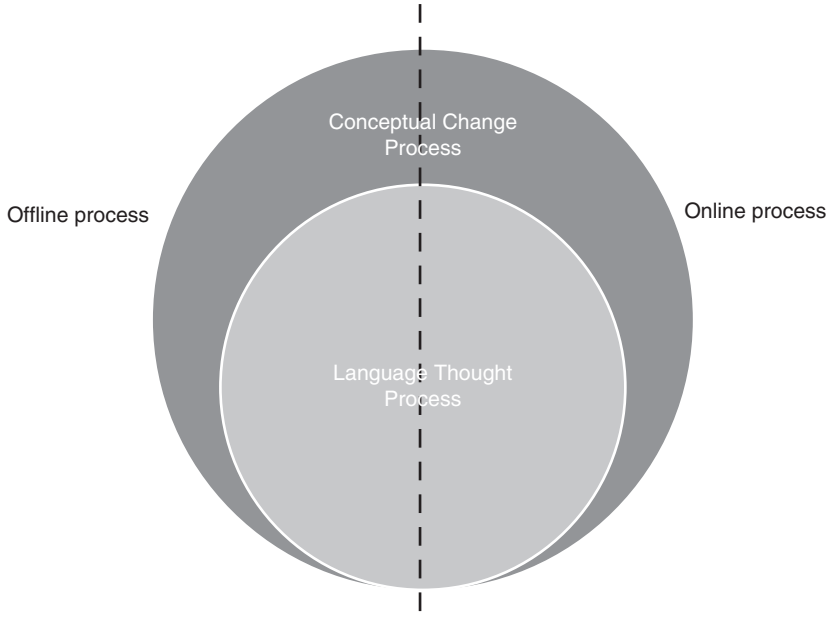

Figure 6.1. Language-thought processes are not a cognitive kind: schematic diagram showing the relation of language-thought processes to processes of conceptual change and indicating their division into offline and online processes.

6.6 Conclusion

Having tried to formulate the LT hypothesis with sufficient precision that it can be evaluated scientifically, it seems likely that it does not identify a single or distinctive cognitive kind. It is not distinctive because the processes that mediate the effects of language on cognition cannot be distinguished in a principled way from a broader class of processes (notably, the lexicalization of novel concepts within a natural language, leading to cognitive restructuring). Moreover, it is not a single kind because these processes do not seem to be homogeneous, since some of them involve covert and simultaneous recruitment of language while others do not, as outlined in the previous section. Perhaps the only thing that sets apart those phenomena regularly classified as LT effects is that they are unexpected and significant cognitive effects involving natural language. Cognitive differences that align with the (somewhat artificial) boundaries between natural languages are clearly of interest to human investigators, particularly since language is commonly thought to be a transparent medium that serves

merely to convey thought rather than to shape or influence it in some way. But such effects may not be cognitively significant in their own right. Does this mean that the whole project of looking for LT effects ought to be abandoned? Not necessarily, but it does suggest that investigators need to look more broadly at cognitive effects involving conceptual misalignments among psychological subjects, whether speakers of different natural languages, or experts and novices, or speakers of different dialects within the same language. Natural language is a convenient marker of possible conceptual disparity, but it is not the only way in which human beings can differ in their conceptual repertoires. Where there are such disparities, there is room for the existence of both "online" and "offline" LT effects, as we have seen. Attributing LT effects to (at least) two distinct cognitive kinds opens the door to a better understanding of these kinds of processes. The natural-kinds approach confirms that there are two different kinds of causal pathways by which language might produce robust effects on human cognition, which is a realization that seems to have emerged gradually among researchers working in this area over the past decade or so. Also, by adopting a natural-kinds approach to the question of the interaction of language and thought, it becomes clearer that there is no principled distinction between LT processes and processes of conceptual change or augmentation.[11] This suggests that researchers need to inquire further into processes that involve cognitive restructuring, to determine the continuities and discontinuities between those involving competence in a natural language (e.g. mastery of color terms, spatial terms, gender terms) and those involving the acquisition of novel concepts in specialized domains, such as scientific theories domains of expertise (e.g. mastery of terms from surfing, wine-tasting, physics, climate science).

[11] Even though there may not be a strictly scientific reason for lumping all LT effects together and treating them as a separate class of cognitive phenomena, it is understandable that we might be especially interested in the cognitive effects that result from speaking different natural languages, for at least two reasons. First, as mentioned above, language is often considered a neutral medium for expressing thoughts rather than a shaper of those thoughts, and hence any evidence showing that natural language has some systematic influence on the thought of language users is of interest. That is particularly the case because the influence has sometimes gone unnoticed (and perhaps because those who have noticed it have tended to exaggerate it). Second, differences between natural languages are important to human beings given the cultural and political associations of natural languages. It is not surprising that any correlation between one's natural language and one's thought processes, albeit small, is of widespread interest both within and outside the scientific community.