

11 ‘My language limits my thoughts’

There’s no denying that speaking and thinking are closely related. We use words to convey ideas to other people, and to organize our own thoughts. We’re so accustomed to giving verbal expression to our thoughts that it’s easy to believe that we ‘think in words’. If language is so central to our thinking, then it seems only logical that the particular *form* of our language – the words that it has, the aspects of the world that it encodes grammatically, and so on – should have a powerful effect on how we think.

The suggestion that language influences thought has a long history, both inside and outside the field of linguistics. The idea is a compelling one; it’s intriguing to contemplate the possibility that our minds are shaped in important ways by accidents of our native language. And not only is it a theoretically interesting idea, but if language really does affect thought, then the real-world consequences could be quite serious. Many people are concerned that politicians, advertisers, and others use language in manipulative and misleading ways. It’s also been claimed that particular languages are well suited to certain purposes because of the kinds of thought they permit or encourage – or, on a more sinister note, that some languages prevent certain kinds of desirable thought, and therefore confine their speakers to a kind of mental straitjacket. If this is true, then maybe there are languages that deserve to be forgotten.

Unfortunately, the true nature of the relationship between language and thought turns out to be very difficult to study, and the history of thinking about this question is littered with questionable assertions, inaccurate claims, and just plain wild speculation presented as fact. But the good news is that a substantial body of rigorous experimental work on the topic has appeared in the last few decades. In this chapter, we’ll look at popular ideas about the relationship between language and thought; then we will examine some relevant experiments, asking what the evidence says about whether, and to what degree, thought is shaped by language.

11.1 The Sapir-Whorf hypothesis

The idea that language influences thought is closely associated with Benjamin Lee Whorf, a linguist who worked in the first half of the twentieth century.

Whorf documented and analyzed several indigenous American languages; at the time, linguists were just starting to grapple with these languages, many of which are different in important ways from more familiar European languages. Their discoveries had profound implications for linguistic theory, in part because they broadened linguists' understanding of what's possible in human language. These differences also invited speculation about the relationship between language and thought: if the grammar of language A seems to divide up the world into concepts that are very different from those of language B, does that mean speakers of language A and language B actually see the world differently?

Whorf thought so. In several of his writings, Whorf discussed words and grammatical constructions in indigenous American languages that have no obvious counterpart in English; he argued that these examples demonstrate different thought patterns. One of his most famous examples involved words for the idea of 'snow': Whorf claimed that Eskimos have many distinct words for 'falling snow, snow on the ground, snow packed hard like ice, slushy snow', and so on, whereas English has only one word for all these things (Whorf 1956, 216). Apparently, what one language takes to be a single concept may be divided up into multiple concepts in another language.

Both Whorf and his mentor, Edward Sapir, argued at times for a close relationship between language and thought; thus, the *Sapir-Whorf hypothesis* is the general idea that the language a person speaks may influence the way that person thinks. We can distinguish between at least two versions of the hypothesis: a *strong* version, whereby your language actually determines and even restricts the way you think; and a *weak* version, whereby your language influences the way you think but doesn't make any idea literally unthinkable. Both Sapir and Whorf argued for stronger and weaker versions of the idea at various times, and it's impossible to know exactly which versions of the hypothesis (if any) they would support if they were still alive today. Even though we credit Sapir and Whorf with the basic suggestion that language and thought are connected, we understand that they wouldn't necessarily endorse every specific proposal that has been put forward under the umbrella of the Sapir-Whorf hypothesis.

Whenever someone claims that some difference between languages A and B causes speakers of those languages to think differently, the first step is to ask whether the languages really are different in the relevant way. One obvious objection to Whorf's Eskimo example, for instance, is that English does *not* have only a single word for the concept 'snow': *snow*, *sleet*, *slush*, *blizzard*, and so on are all words that refer to snow in various states; skiers and other winter athletes have even more precise terms. (Incidentally, Whorf's snow illustration is also at least partly responsible for the modern urban legend that 'Eskimos have hundreds of words for snow'; see Pullum 1991 for an entertaining discussion of the myth.)

But there are other cases where languages really *do* describe the world in different ways. English, for example, has a single basic color term *blue* that covers a range of hues, from the color of the sky to the color of blueberries. Russian, by contrast, has two words: *goluboy* covers lighter blues, and *siniy* darker blues. English speakers can refer to *light blue*, *dark blue*, *sky blue*, *navy blue*, and so on – but these aren't basic terms the way *goluboy* and *siniy* are. Conversely, Russian lacks a single word that covers the entire range of English *blue*. In this case, the linguistic difference is genuine.

But it's not enough to establish that languages A and B really are different; we also have to be certain that speakers of those languages actually *think* differently, and we must do so in a way that avoids simply restating the fact that the two languages are different. It would be very boring, for example, to observe that Russian speakers describe shades of blue with two different words where English speakers use only one – this is just another way of saying that Russian speakers speak Russian and English speakers speak English. What we need is evidence that Russian and English speakers actually *think* about blues differently, independent of the way they talk about them. One criticism of Whorf's writing is that he often failed to do this. Although he sometimes argued for distinct habits of thought based on larger cultural patterns (see, e.g., Whorf 1941), he also sometimes fell into the trap of proclaiming that two groups obviously think about things differently just because they use different words – the snow example is of this type. To be fair, this is true of most discussions of language and thought until the last decades of the twentieth century; studying thought independent of language turns out to be very difficult to do. But we can't accept the Sapir-Whorf hypothesis without good evidence, and we should not abandon the search for good evidence just because it's hard.

Finally, even if we can show that speakers of two languages talk *and* think about something in different ways, that's not enough to prove that their language is affecting how they think. We also have to show that the linguistic differences are *causing* the differences in thought. Take the example of languages that have only a few number words – e.g., words for 'one', 'two', and 'three', and another term meaning 'many' that covers larger quantities. If we find that speakers of a language like this have trouble with mathematical concepts involving larger numbers, is that because their language prevents them from thinking about exact numbers larger than three? Or is it because their society has no need for counting up to larger numbers, which explains *both* speakers' difficulty with the concept (they've never practiced) *and* the lack of number words (speakers simply don't need them)?

The Sapir-Whorf hypothesis, then, turns out to be extremely difficult to test. At the end of this chapter, we will examine some of the creative techniques

researchers have come up with to explore the relationship between language and thought. First, though, we will survey some popular Whorfian beliefs and ask how plausible they are.

11.1.1 *Reasons to doubt the strong version of the Sapir-Whorf hypothesis*

In its strongest form, the Sapir-Whorf hypothesis suggests that I'm literally unable to think about things that can't be expressed in my native language. There are many common-sense reasons to believe that this strong form of the hypothesis simply can't be true. It would be false, for example, to suggest that English speakers can't distinguish among various shades of blue (because they have only a single word *blue*) – English speakers obviously *can* do this, and if they need to refer to a specific type of blue, they will find a way to do so (*light blue*, *pale blue*, *sky blue*, *the color of that guy's shirt*, and so on).

If the strong version of the Sapir-Whorf hypothesis were true, it would be impossible to invent new things or imagine new ideas: if my language doesn't already have a word for the thing, I should be unable to think about it. But of course people come up with new ideas all the time; if they lack a word for the thing, they either create or borrow one. The words *telephone* and *computer* didn't exist in the 1500s, but this didn't stop people from inventing those things and naming them. Similarly, ideas have no problem spreading across linguistic boundaries: it's easy to think of political ideas, religious concepts, and even particular types of food or dress that have been borrowed across language divides. Again, there's no reason to think that people are prevented from thinking about new ideas simply because they happen to lack the relevant word. If my language has no word for the hairstyle known as a *mullet*, and I've never seen one before, it's preposterous to think that I would be unable to understand what I was looking at if you showed me an example, or that I would be unable to describe it to others. If mullets became common enough in my daily life that I needed a single word to refer to them, I would either borrow your word or create a new one.

Indeed, under the strong version of the Sapir-Whorf hypothesis, it's a mystery how we ever learn words at all. Babies are obviously capable of thought long before they learn language: they can recognize familiar faces before they know (or recognize) words like *Mom* and *Dad*; they learn that they can pull on certain kinds of handles to discover interesting things before they learn the word *drawer*; they know how to operate toys and generalize concepts such as 'buttons' or 'lids' before they are even remotely close to being able to describe what they're doing. Under the strong version of the hypothesis, much of what happens in formal education ought to be impossible because it involves a catch-22: to tell you what a new word like *logarithm* or *extradition* means, I have to describe the concept to you, but you won't be able to understand the concept

unless you already have a word for it. Similarly, the strongest version of the hypothesis predicts that it should be impossible to learn a second language, or at least impossible to learn those words that don't precisely correspond to words in our first language.¹

The strong version of the Sapir-Whorf hypothesis assumes that thinking depends on words, but there are plenty of reasons to believe that this isn't necessarily the case. Everyone has had the experience of having a word on the 'tip of your tongue' – you have the concept clearly in mind, and you know that you know a word for it, but you simply can't remember what the word is. Even more tellingly, we can have a concept without having a word to go with it. For example, English has standard names for each of the fingers (*thumb*, *index finger*, and so on), but on the foot, only the *big toe* and the *little toe* have special names. However, the fact that I don't have specific names for my other toes doesn't mean I can't distinguish among them, or that I couldn't describe them to someone else – if pressed, I might refer to my *second toe* or my *index toe* or *the toe next to my big toe*.² There are also plenty of people who don't have language but clearly have some form of thought. Babies, as mentioned above, are an obvious example; we could add to the list adults who have conditions that interfere with language, such as severe autism or aphasia. We can also identify aspects of our ordinary thinking that don't seem to involve language at all: imagining music, for example, or picturing a visual image.

All these examples raise serious doubts about the strong version of the Sapir-Whorf hypothesis. We are apparently not slaves to the words we use, even if it turns out that we are more subtly affected by them. Thought and language are, at least to some degree, independent.

11.1.2 Popular beliefs that language influences thought

For all the reasons just discussed, most linguists reject the strong form of the Sapir-Whorf hypothesis. Lucy (1996) takes issue with the fact that linguists typically begin discussions of the hypothesis with an argument against its strong form, much like the one above. He suggests that the strong version is a straw man: since most researchers are actually investigating much weaker forms of the hypothesis, it's unfair to spend so much time discrediting the

¹ This is not to say that anyone from any group can understand any concept. There really are societies, for example, whose members have extreme difficulty with basic mathematical operations. It's also common for people who move from one society to another to have trouble understanding the values and practices of the new culture in which they find themselves. The point is that before we attribute these difficulties to *linguistic* differences, we also have to ask whether they could be due to differences in culture or experience.

² Anatomists, of course, have standard names for the toes as well as for other parts of the body – e.g., the *philtrum*, which is the groove between the nose and upper lip – but these aren't widely known by non-specialists, who nevertheless know that these things exist.

strong form. Lucy may be right about linguists, but outside the ivory tower (or even in academic disciplines other than linguistics), it's surprisingly common to find strong Whorfian ideas about the connection between language and thought.

Someone learning a foreign language, for example, will inevitably discover concepts that the new language encodes differently from his native one. Many people find these differences intriguing, since they point to alternative ways of talking about the world. Combine this with the fact that many people study foreign languages and culture together, and you have a recipe for widespread belief that linguistic differences are linked to differences in thought.

One way languages can differ, of course, is in the particular words they have; it's common knowledge that a given word may have no exact translation from one language into another. Two languages may have words that cover the same range of meaning but divide it differently, as in Russian *goluboy* and *siniy* vs. English *blue*. Alternatively, one language may have a single word for a concept with no exact counterpart in another language; a commonly cited example is German *Schadenfreude*, a word that refers to taking pleasure in someone else's misfortune and has no exact English translation.

What do these differences mean? One conclusion might be that if a language has no word for such-and-such a concept, it must be because speakers of that language don't have the concept itself – indeed, maybe they can't even conceive of the idea. (A collection of claims along these lines in the popular media can be found in Liberman 2009.) Sometimes this is framed as a positive thing: that some language, for example, has no word for 'lying' because its speakers are always honest. But the supposed lack of some concept can also be presented as a moral failing:

A particularly damaging example of the *No word for X* fallacy is one that one hears here in Northwestern Canada. Many of the Athabaskan languages of Canada have a word for "thank you" that is borrowed from French *merci*. . . . This fact has suggested to the ignorant that these languages previously had no word for "thank you", from which they draw the further conclusion that their speakers had no concept of gratitude. Such a people, of course, must have been sub-human savages. The conclusion is that it's a good thing that white people came to rescue them from their degraded traditional way of life.

Poser (2006)

One problem with this kind of claim is that it's usually false: Athabaskan languages, for example, have plenty of native ways of expressing thanks. But even when the claim is true, we've already seen that lacking a word for something doesn't make a person incapable of thinking about that idea. Speakers will find a way to say what they need to say – they will either come up with a phrase that describes the relevant concept, or they will borrow or invent a

new word. In the case of *Schadenfreude*, English speakers have happily borrowed the word from German, with the result that it's now false to say that English has no word for taking pleasure in someone else's misfortune. The upshot is that examples of words without precise equivalents in other languages make nice factoids, but they don't tell us much about deep cultural or cognitive differences.

Languages also differ in the kinds of meaning that are marked grammatically. Whorf (1941, 1956) observed that Hopi doesn't make the same past-present-future distinction found in many European languages, and concluded that speakers of Hopi don't have the same concept of time as Europeans. (The facts of Hopi are still somewhat unclear, but there are certainly other languages, such as Mandarin, that are widely agreed to lack grammatical tense.) In English, many nouns are obligatorily marked for number: *I saw the cat* means I saw exactly one, while *I saw the cats* means I saw two or more. By contrast, some languages have an additional dual marker (i.e., a form meaning 'exactly two'), while others have no grammatical plural at all. Languages such as Turkish (but not English) have grammatical markers, known as *evidentials*, by which the speaker indicates the source of his knowledge about what he is saying: if I personally witnessed Ahmet arrive, I can say *Ahmet gel-di* 'Ahmet came'; but if I only heard about it, I have to say *Ahmet gel-miş*.

Once again, the mere fact that languages can differ in these ways is no reason to jump to strong Whorfian conclusions. The Hopi example has led to the popular belief that Hopi speakers have no sense of time whatsoever, an idea that is contradicted by the fact that the language has plenty of other resources for talking about time. Mandarin speakers, with no grammaticalized plural marker, are nevertheless perfectly capable of telling the difference between one object and more than one. English speakers don't have grammatical evidential markers in their language, but they're perfectly able to describe the source of their information if necessary (compare *John is single again* with *Apparently John is single again*).

Even 'exotic' differences among languages, then, are no reason to embrace the strong version of the Sapir-Whorf hypothesis. But this doesn't necessarily mean that these differences don't matter at all. In recent decades, linguists have started to explore the possibility that grammatical categories can direct speakers' attention in subtle but real ways. The idea is that an English speaker, for example, is constantly forced to pay attention to the difference between single things and multiple things because her language requires her to make this distinction. A Mandarin speaker, on the other hand, is perfectly capable of noticing this difference but isn't forced to do so every time she wants to talk; therefore, maybe Mandarin speakers are less attentive to number differences than English speakers. We will see some examples below of studies that test this kind of weak version of the Sapir-Whorf hypothesis.

One last reason many people find the Sapir-Whorf hypothesis appealing is that it sometimes resonates with their experience of using more than one language. Many bilinguals feel that they think differently when they use different languages, or even that they become different people. On the one hand, reports like these reflect real experiences, and therefore they're something we have to explain if we want to understand language in all its uses. On the other hand, we can't assume too quickly that these experiences are due solely to language differences, as opposed to other factors. An English-Japanese bilingual may feel different in each language, not because the languages themselves are different, but because the act of speaking English brings up one set of associations (with particular people, situations, and cultural practices) while the act of speaking Japanese brings up another. In addition, many bilinguals report other experiences that are actually incompatible with strong versions of the Sapir-Whorf hypothesis. For example, someone may be able to remember the content of a conversation that he has had, but not the language in which it was conducted. Under strong versions of the hypothesis, this should be impossible – if thoughts are determined by words, how can I remember the thoughts I had but not the words I used to express them?

11.2 George Orwell and political language

11.2.1 *Newspeak*

We see a slightly different flavor of the Sapir-Whorf hypothesis when we consider politics; politicians are widely perceived as using language to manipulate and fool the masses. The basic idea is that unpopular or inconvenient facts are given nice-sounding names to disguise their true nature; this obfuscation fools the public into supporting policies, people, and institutions they would otherwise reject.

Popular discussion of these issues has been enormously influenced by the writings of George Orwell. In one well-known essay, 'Politics and the English language', Orwell identifies several specific examples of pleasant expressions being used in his own day to cover up unpleasant realities:

Defenceless villages are bombarded from the air, the inhabitants driven out into the countryside, the cattle machine-gunned, the huts set on fire with incendiary bullets: this is called *pacification*. Millions of peasants are robbed of their farms and sent trudging along the roads with no more than they can carry: this is called *transfer of population* or *rectification of frontiers*. People are imprisoned for years without trial, or shot in the back of the neck or sent to die of scurvy in Arctic lumber camps: this is called *elimination of unreliable elements*.

Orwell (1968, 136)

Orwell's argument in this essay is that politicians use vague language to hide what they're actually doing. Everyone, he insists, should use language in a way that is clear, precise, and concrete.

An even more sinister view of political uses of language appears in Orwell's work of dystopian fiction, *1984*. This novel depicts a totalitarian government that has unprecedented control over every aspect of citizens' lives; leaders demand absolute obedience in action and even in thought. The appendix to the novel describes the government's long-term plan for enforcing compliance: a language called *Newspeak*.

The purpose of Newspeak was not only to provide a medium of expression for the world-view and mental habits proper to the devotees of Ingsoc, but to make all other modes of thought impossible. It was intended that when Newspeak had been adopted once and for all and Oldspeak forgotten, a heretical thought – that is, a thought diverging from the principles of Ingsoc – should be literally unthinkable, at least so far as thought is dependent on words.

Orwell (1949)

This goal was to be accomplished by eliminating words such as *justice* or *democracy* that were incompatible with the ruling party's ideology. Without access to words like these, citizens would be unable to think about the relevant ideas, and thus would be unable to desire or demand them. Note that Newspeak could succeed only under a very strong version of the Sapir-Whorf hypothesis: its crucial assumption is that not having a word for something means not being able to think about that thing.

Orwell's writings have been widely hailed as prophetic, and words such as *Newspeak* and *Orwellian* have entered the general vocabulary. A quick search of major newspapers, blogs, or other venues for public writing turns up plenty of complaints about this or that politician's use of language, often explicitly invoking Orwell or accusing the target of Newspeak. The fear is that if the public passively accepts politicians' misleading descriptions, they'll end up acquiescing to policies that are obviously wrong; the situation could be avoided if only politicians and the media were forced to use clear, neutral language.

11.2.2 Real-world examples: Can misleading language disguise the truth?

How frightened should we be? Are we really in danger of becoming slaves to manipulative political language? There are several reasons to think the situation isn't quite so dire. The first is the fact that Newspeak depends on such a strong version of the Sapir-Whorf hypothesis – but, as we have already seen, it's highly unlikely that thought really does depend totally on language.

Another reason to suspect that even manipulative language can't fool everyone is the very fact that we're able to talk about it. Surely Orwell didn't think it was impossible to see the real meaning of terms like *pacification*; after all, Orwell himself managed to do it. The very act of pointing out a term that you find offensively misleading, such as *collateral damage* or *rev-enue enhancement*, demonstrates that at least one person (namely, you) wasn't fooled.

Indeed, when we look at real-world examples, we see that even nice-sounding phrases can easily acquire negative associations. Consider some examples from the United States:

- Some institutions give favorable treatment or consideration to members of minority groups; policies like these are known as *affirmative action*. The positive word *affirmative* hasn't prevented some people from concluding that such policies are unjust; the result has been a vigorous, decades-long debate over whether affirmative action is a good or bad thing. For people who disagree with these policies, *affirmative action* is a distinctly negative term.
- Shortly after the terrorist attacks of September 11, 2001, Congress passed a law that expanded the authority of law enforcement agencies. This law is known as the *Patriot Act* – again, a vaguely positive term obviously meant to encourage support for the law. Critics, though, have argued that the law is intrusive and violates civil liberties. *Patriot Act* is a highly negative term among people who disagree with the law.
- On a lighter note, the halftime show of the 2004 Superbowl became notorious for a performance by singers Justin Timberlake and Janet Jackson: Timberlake uncovered part of Jackson's costume, exposing one of her breasts on live television. The ensuing controversy was intense, with commentators debating whether the exposure was intentional, whether anyone deserved to be penalized, and whether the incident was newsworthy at all. There was one thing, though, that everyone could agree on: when Timberlake referred to the incident as a *wardrobe malfunction*, the term met with instant and universal ridicule. The euphemistic language didn't cause the public to view the incident more mildly; rather, the public's knowledge of what had really happened made the euphemistic language risible.

Authoritarian governments in the real world, unlike the one in 1984, have had little success in using Newspeak-like language to prevent their citizens from recognizing what is plainly before their eyes. Kershaw (1983, 194–199) observes that when ordinary Germans saw that Nazi propaganda conflicted

with the reality on the ground, their reaction was to disbelieve nearly everything the government said – in fact, radio broadcasts from Great Britain became more trusted than the German government's own pronouncements. Cameron (1995, 152–155) cites an unpublished paper by Julian Konstantinov describing how the collapse of Communism in Bulgaria revealed Bulgarians' real attitudes toward the words they had been forced to use: citizens immediately abandoned terms such as *People's Army* or *People's Republic*, for example, because they observed that these institutions served not the people but the party. In other words, the meaning of *People's* had changed so that it corresponded to reality; the meaning intended by those in power (that these institutions reflected the will of the people) collapsed under the weight of fact.

Contemporary examples of similar phenomena can be found in modern China. Public speech is tightly controlled, particularly on the Internet, and one of the official justifications for restrictions on free speech is the need to promote a 'harmonious society'. Some citizens have responded by adopting the official language but giving it a subversive meaning: a blog post or online comment can be *harmonized*, which simply means that it's been censored. Again, people's perception of reality doesn't change just because they have to use euphemistic words; rather, the meanings of the words change so that they match reality.³

Finally, Cameron (1995, 72–75) notes that Orwell's goal of clear and unbiased language may be impossible anyway. The problem is that we use language *both* to describe 'the plain facts of the matter' *and* to show what we think about those facts. Consider, for example, an opponent of Newspeak who argues that every action that involves ending a human life should be described as *murder*. On the face of it, this sounds like a sensible proposal. But things become much more complicated when we consider practices that are controversial precisely because people disagree over whether they fall into the moral category of murder. It's easy to find opinion pieces, for example, that explicitly invoke Newspeak in their criticism of abortion (Broomstreet 2006), euthanasia (Hentoff 1987), or capital punishment (Aronson 2007). No one disagrees about the relevant facts (namely, that all of these actions result in death); rather, it's our *evaluation* of those facts that is at issue.

Note that all these examples involve a population that knows the truth of the matter, and is therefore able to notice the difference between the facts and what's said about them. When people do *not* know the facts, then of course the

³ Another popular technique is to replace potentially sensitive words with homophones or near-homophones. 'River crabs' have become a symbol of censorship because, in Mandarin, *river crab* sounds like *harmony*. River crabs are said to threaten the beautiful but endangered *grass-mud horse* – a fictional creature whose name sounds like an obscene phrase, and which symbolizes resistance to censorship. The *China Digital Times* (<http://www.chinadigitaltimes.net/>) maintains a list of phrases like these in its 'Grass-mud horse lexicon'.

situation is different, and language can easily be used to disguise and mislead. When a politician falsely declares that she's never taken a bribe, her constituents may well believe her if they have no reason to think otherwise. But we can understand this phenomenon without ascribing any magical powers to language; it's an ordinary part of our everyday experience, called *lying*.

The real world, then, gives us little reason to believe that large-scale linguistic manipulations like Newspeak will become reality anytime soon. This is not to say, though, that the words we use don't matter at all. As noted above, we use language for taking positions as well as describing facts, and the world has no shortage of issues that are well worth discussing. In addition, we use language as a framing device, and there's evidence that different ways of framing a situation really matter (see section 11.3.5). But language isn't the only tool we use to influence people; this is why executives wear suits, negative political ads feature ominous music, and consumer products are covered with pictures of smiling faces. We can appreciate the framing potential of language without concluding that language is unique, or that we're slaves to the words we hear.

11.3 Case study: Does our language affect the way we think?

The strong version of the Sapir-Whorf hypothesis seems highly improbable – our thoughts aren't completely determined by the language we use. But the weak version, the idea that language has a subtle influence on the way we think, could very well be true. As discussed above, the weak version of the hypothesis is difficult to test because we have to find a way to assess thought independent of language. In this section, we will look at several experiments designed to do just that. As we do so, we'll ask whether they provide us with good evidence for a link between language and thought, and if so, what the nature of that link is.

11.3.1 *Color terms: Winawer et al. (2007)*

For several decades during the early and middle twentieth century, variants on the Sapir-Whorf hypothesis were very popular. The domain of color terms, in particular, seemed to confirm that languages (and therefore, presumably, thought) could differ in large and unexpected ways. Linguists and anthropologists observed that languages divide up the color spectrum differently: some languages, for example, have separate basic color terms for 'light blue' and 'dark blue'; others, such as Japanese, have a single basic term that describes both 'blue' and 'green'. Many languages have no word at all for colors that English speakers would consider basic, such as pink or grey. Some languages have only a tiny number of basic color terms, as few as two or three. This range

of variation suggested to many people that whether or not a person considers two colors to be ‘the same’ will depend on whether they are called by the same name in that person’s language.

Received opinion about the implications of these differences began to change in the 1960s and 1970s; many researchers came to accept the idea that the variation they saw actually reflected fundamental similarities in human cognition across cultures. Several factors contributed to this shift. One was the growing influence of Noam Chomsky, who argued that humans are born with much of their knowledge of language ‘pre-programmed’, and that differences in grammatical structure across language are actually just variations on a universal core. Chomsky’s proposals aren’t actually incompatible with the Sapir-Whorf hypothesis – it could be that even those differences end up affecting thought in profound ways – but in practice, linguists who are sympathetic to Chomsky’s ideas tend to be skeptical of Whorfianism, and vice versa.

Another factor that contributed to the shift towards a universalist view of color was Berlin and Kay’s (1969) influential survey of color terms in 98 languages. Berlin and Kay argued that there are actually strict constraints on the color terms in a given language: if a language has n basic color terms, there are only a few possibilities for what those terms will be. They proposed a hierarchy of color terms based on when they appear in a language’s vocabulary: a language with only two color terms will have words for ‘dark’ and ‘light’; a language with three terms will additionally have a word for ‘red’; and so on. The observation that color terminology doesn’t vary without limit suggested to many linguists that the color spectrum can’t be divided up in arbitrary ways; there must be innate constraints on how color is perceived that can’t be overridden by language.

Finally, Heider (1972) and Heider and Olivier (1972) conducted a series of experiments with the Dani people of New Guinea, a group with only two basic color terms: *mola* (‘light’) and *mili* (‘dark’). Dani and English speakers actually performed very similarly on some tasks, such as remembering colors or learning new names for colors, despite the differences in their native color vocabulary. Heider argued that there are certain colors – ‘focal colors’, such as a prototypical red – that are inherently salient, and people are much better at identifying and remembering focal colors than non-focal colors. By this argument, a language doesn’t make its speakers better at seeing some colors than others; rather, some colors are inherently more noticeable, and languages tend to have basic color terms that cluster around those focal points. Not every language has a word for every focal color, but speakers will be sensitive to focal colors regardless of whether they happen to have words for them.

Studies like these, combined with the changing Zeitgeist, convinced many linguists in the second half of the twentieth century that the effect of a person’s native color terms on her perception of color is weak or non-existent. But in

recent decades, a number of researchers have begun to argue that color perception really *is* influenced by a person's native vocabulary, albeit in subtle ways. Roberson et al. (2000), for example, replicated several of Heider's experiments with a different group and found that native language did affect color perception. Other researchers have explored how people perceive hues that lie at the boundary between one color term and another – for example, how English speakers perceive colors that are very close to the boundary between *blue* and *green*.

In one such study, Winawer et al. (2007) investigated the phenomenon known as *categorical perception*, in which a person is more sensitive to differences across categories than between categories. English speakers, for example, are known to perceive the difference between *blue* and *green* categorically: they do well at distinguishing between two colors if one is 'blue' and the other is 'green', but they're much worse at distinguishing between equally distinct colors if both are 'blue' or both are 'green'.

If a language's color terms cause speakers to perceive a category boundary wherever they have a linguistic boundary, then we might expect Russian speakers to perceive the difference between light and dark blue (*goluboy* and *siniy*) categorically where English speakers do not. Winawer et al. studied 26 native speakers of Russian and 24 native speakers of English. During the experiment, each subject would see three blue squares on a computer screen, one at the top and two at the bottom; the subject's job was to identify which of the two bottom squares exactly matched the shade of the top square.

Russian and English speakers did in fact behave differently. Russian speakers responded faster when the two bottom squares straddled the boundary between *goluboy* and *siniy*; in other words, they performed better when they were distinguishing between colors that belonged to two different categories. They responded more slowly when the two bottom squares were both *goluboy* or both *siniy*. English speakers, by contrast, responded equally quickly regardless of whether the two squares were both *light blue* or *dark blue*, or whether one was *light blue* and the other *dark blue*. Thus, Russian speakers showed evidence of a category boundary in the middle of the blue spectrum, but English speakers didn't.

So far, these results look promising. Moreover, Winawer et al. found these differences in a non-linguistic task – that is, subjects were apparently just matching colors, not describing them verbally. But it turns out that things aren't quite so simple. Although subjects weren't *required* to name the colors during the color-matching task, they might have done so anyway. Maybe, for example, the Russian subjects were covertly saying *goluboy* and *siniy* to themselves when they saw the colored squares, and if they found that one of the bottom two squares was *goluboy* while the other was *siniy*, that linguistic label gave them an extra boost in matching them to the top square. In other

words, maybe language differences helped the Russian speakers only because they were actively using language to perform the task, not because their perception had been permanently altered by their language. It could even be that Russian speakers were naming the colors subconsciously, not fully aware of what they were doing. This issue turns out to be a serious and recurring problem for studies of language and thought: even if researchers don't explicitly ask subjects to use their language to accomplish some task, subjects may invoke language anyway. The result is that even differences that appear to be purely non-linguistic may turn out to be linguistic after all. Papafragou et al. (2002, 216) sum up the dilemma by concluding that 'many apparent effects of language on thought are more appropriately interpreted as effects of language on language.'

Winawer et al. addressed this possibility by conducting two more versions of the same experiment, which they made harder by asking subjects to perform an extra task at the same time that they were matching the colored squares. In one of these interference conditions, subjects were given an eight-digit number and asked to recite it to themselves silently; after several color-matching trials, they had to recall the number. The idea was that if subjects were performing a demanding verbal task (reciting a number), they couldn't use language to help them in the color-matching task, and the category advantage for Russian speakers would disappear. In the other interference condition, subjects saw a grid of black and white squares; they had to hold a picture of the grid in mind for several color-matching trials and then recall it. The idea was that because this task is demanding but doesn't involve language, subjects would be free to use language for the simultaneous color-matching task, and the category advantage for Russian speakers would remain.

Figure 11.1 shows the results for all three types of interference (none, spatial, and verbal). The predictions are confirmed perfectly: Russian speakers maintained their advantage for cross-category comparisons when they were performing the spatial interference task, but the advantage disappeared in the verbal interference task. English speakers made no distinction between cross-category and within-category comparisons, regardless of the type of interference (or lack thereof).

So, do the differences between Russian and English affect thought, or not? The answer appears to be both yes and no. On the one hand, Russian speakers were slightly better at distinguishing between *goluboy* and *sinii*, although the effect sped up Russian speakers' inter-category responses by only about a tenth of a second. On the other hand, this effect seems to depend crucially on subjects' ability to use language (perhaps unconsciously) to perform the task. In other words, having both *goluboy* and *sinii* doesn't permanently alter Russian speakers' color perception; it merely gives them a slight advantage in some situations to have rapid access to simple color terms that distinguish the two

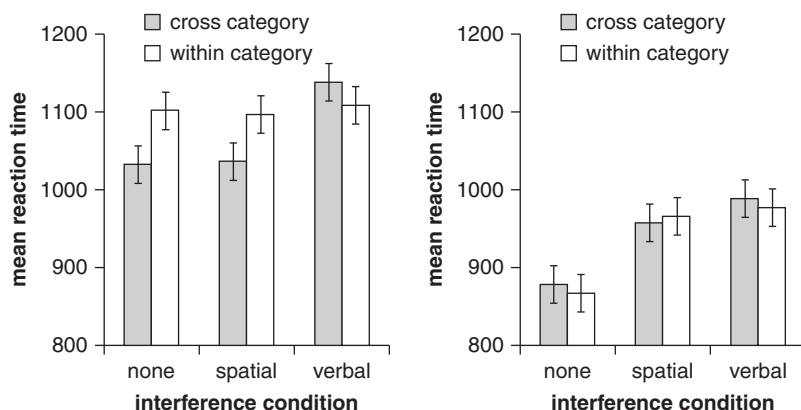


Figure 11.1 Response times (in milliseconds) for across- and within-category comparisons for Russian speakers (left) and English speakers (right) by interference condition. Jonathan Winawer, Nathan Witthoft, Michael C. Frank, Lisa Wu, Alex R. Wade, and Lera Boroditsky, Russian blues reveal effects of language on color discrimination, *Proceedings of the National Academy of Sciences* 104(19):7780–7785, Figure 2. Copyright (2007) National Academy of Sciences, U.S.A.

types of blue. What we have here is evidence for a weak Whorfian effect of language.

11.3.2 Grammatical gender: Konishi (1993)

In languages with grammatical gender, each noun belongs to a specific sub-class – its gender – that affects its grammatical behavior. Often, other words in the sentence are required to carry information about the gender of a particular noun. The adjective modifying a noun, for example, might be required to match the gender of the noun; in some languages, the verb indicates the gender of its subject or object. Thus, a person who speaks a language with grammatical gender has to keep track of which sub-class each noun belongs to.

The basis of these groupings can be more or less arbitrary. In some indigenous American languages, the gender of a noun is based on criteria such as the shape of the thing it refers to (round and compact versus long and thin, for example) or whether or not it's human. In many European languages, by contrast, the gender of a noun has little or no connection to its meaning; the gender is either completely arbitrary or linked to some morphological property of the noun (e.g., Spanish words ending in *-ción* are typically feminine). The interesting thing about gender in European languages is that noun classes are linked to biological sex: the forms used with one class of nouns ('masculines') are

also used for male humans; the forms used with the other class of nouns ('feminines') are also used for female humans. The result is that speakers of such languages regularly talk about inanimate objects with the same grammatical forms that they also use for distinguishing humans according to biological sex.

No one would argue, of course, that speakers of languages with sex-based gender systems literally believe that inanimate objects are male or female. The Spanish word for 'table' (*mesa*) is feminine, but Spanish speakers are perfectly aware that tables don't have ovaries. Mark Twain famously noted that the German word for 'girl', *Mädchen*, is grammatically neuter – neither masculine nor feminine. But no one thinks German speakers don't know the difference between boys and girls.

It's possible, though, that gender influences speakers in more subtle ways. Maybe Spanish speakers, for example, vaguely associate tables with femaleness because of their grammatical gender; Konishi (1993) tested this idea. He asked 40 native speakers each of Spanish and German to rate 32 words along several dimensions such as 'weak'–'strong'. These dimensions were grouped into larger categories related to ideas that are known to be associated with biological sex; for example, the 'potency' category (which included dimensions such as 'weak'–'strong' and 'small'–'big') is associated with sex, with males perceived as more potent ('strong', 'big', etc.) than females.

Crucially, many of the words in Konishi's study have different genders in Spanish and German. Some, such as *punte* and *Brücke* (both meaning 'bridge'), involve a masculine Spanish word and a feminine German word; others, such as *manzana* and *Apfel* ('apple'), involve a feminine Spanish word and a masculine German one. If grammatical gender affects how speakers think, we might expect speakers of each language to rate masculine words as more potent and feminine words as less so, regardless of their meanings.

Figure 11.2 shows that this is exactly what happened: within each language, masculine nouns were rated as more potent than feminine nouns, even though the masculine nouns in one language corresponded to the feminine nouns in the other and vice versa. Konishi reports that the effect was statistically significant at $p < .01$. These differences suggest that speakers' ideas about these words were indeed influenced by the words' grammatical gender: objects described by masculine nouns tended to be thought of as slightly bigger, stronger, etc. than objects described by feminine nouns. Notice, though, that the effect is tiny – just a tenth of a point on a seven-point scale. Clearly, grammatical gender is not causing speakers of these languages to see some objects as entirely masculine and others as entirely feminine; the difference may be real, but it's also extremely subtle.

We might also ask how much these results are mediated by language. Konishi's subjects were instructed to rate 'concepts', not 'words', but the fact remains that they saw the actual words during the experiment. Maybe the

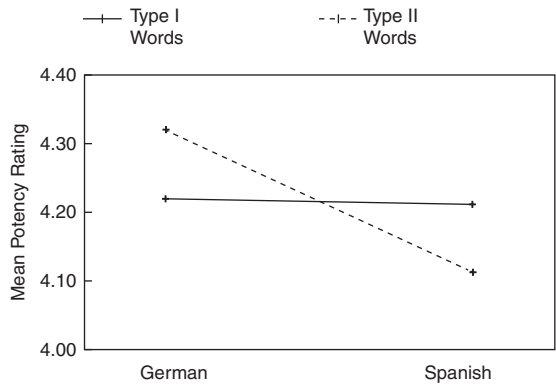


Figure 11.2 Mean potency rating of German and Spanish Type I and Type II words. *Note:* Type I words: feminine gender in German and masculine gender in Spanish (e.g., *sun, fork, pan, brush*); Type II words: masculine gender in German and feminine gender in Spanish (e.g., *moon, spoon, pot, broom*). Toshi Konishi, The semantics of grammatical gender: A cross-cultural study, *Journal of Psycholinguistic Research* 22(5):519–534, 1993. With kind permission from Springer Science and Business Media.

experience of seeing the word *manzana*, complete with the morphological cue to its gender (the final *-a*), brings up feminine associations that don’t arise in the experience of seeing an actual apple. In other words, maybe subjects associate gender with the word, but not the thing.

Overall, the results of Konishi (1993) suggest a weak version of the Sapir-Whorf hypothesis: grammatical gender may indeed give inanimate objects subtle associations with biological sex that they otherwise wouldn’t have. But the effects are small, and gender is only one of the many factors that affect how these objects are perceived. Grammatical gender does *not* overwhelm everything else people know about apples and bridges.

11.3.3 Motion events: Papafragou et al. (2002)

Movement can be encoded linguistically in different ways. Compare, for example, the sentences in (1) and (2) (Papafragou et al. 2002, 195):

- (1)

The man

walked

across

the street.

FIGURE

MOTION+MANNER

PATH

GROUND
- (2)

O andras

dieshise

to dromo

(me ta podia/perpatontas).

the man

crossed

the street

(on foot/walking)

FIGURE

MOTION+PATH

GROUND

(MANNER)

Both of these sentences convey the same information: that the man was in motion, that his movement began on one side of the street and ended on the other side, and that he accomplished this movement by walking. The sentences are different in how they use specific grammatical structures to encode various parts of the motion event. In the English sentence in (1), the main verb (*walked*) encodes both the fact that the man was moving and the MANNER in which he moved (namely, walking). The PATH that he took is encoded in the following prepositional phrase (*across the street*). The Greek sentence in (2) is different. Here, the main verb (*dieshise*) encodes the man's PATH (namely, that he was moving across something – in this case, the street). The MANNER of the motion is expressed in a separate phrase (*me ta podia* or *perpatontas*).

Talmy (1985) first observed that languages differ systematically in how they encode motion events. In some languages, like English, motion verbs typically encode the MANNER of the movement but not the PATH: *walk*, *crawl*, *float*, *roll*, *bounce*, and so on. (English does have verbs that encode PATH rather than MANNER, but they're less common: *cross*, *ascend*, etc.) Other languages, like Greek, have the opposite tendency: motion verbs typically encode the PATH of the movement, and MANNER (if expressed at all) requires a separate phrase. Note that these differences affect the type of information that has to be specified in a sentence that describes motion. If the verb encodes the MANNER of motion, it's possible to produce a sentence that doesn't mention the PATH at all: *The man walked*. Conversely, if the verb encodes information about the PATH, it's possible to produce a sentence that doesn't mention the MANNER of motion: *The man crossed the street* (but he could have been walking, running, crawling, flying, etc.).

What we have, then, are not differences in what English or Greek *can* express, but rather differences in what they *usually* express. A number of studies (e.g., Slobin 1991) have confirmed that speakers of English are far more likely to use MANNER verbs when describing motion events, while speakers of Spanish (a 'path' language like Greek) are far more likely to use PATH verbs. Could it be that English speakers actually pay more attention to the MANNER of motion events, while Spanish or Greek speakers pay more attention to PATH?

Papafragou et al. (2002) studied speakers of English and Greek from a wide range of ages (four years old to fifty) to test whether the two groups actually think about events differently. They showed each subject a series of drawings that depicted people or animals in motion and asked subjects to describe what they saw; not surprisingly, Greek speakers were more likely than English speakers to use PATH verbs (e.g., *The frog is entering the room [jumping]*), while English speakers were more likely to use MANNER verbs (e.g., *The frog is jumping [into the room]*). Two days later, the researchers showed more pictures to the same subjects and asked them to remember whether these were

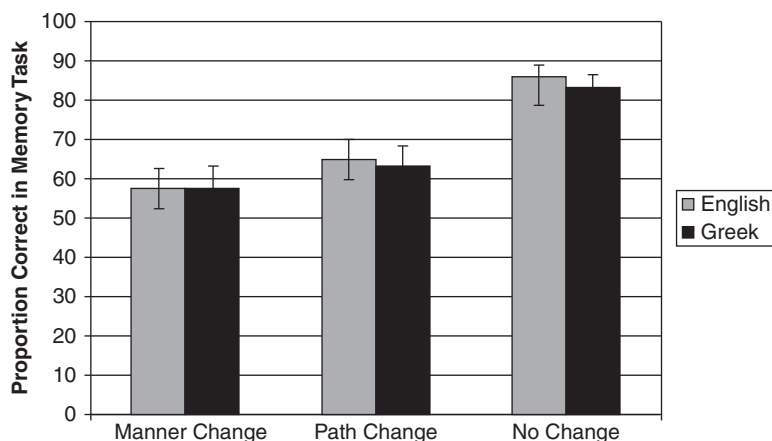


Figure 11.3 Performance on the memory task of Experiment 1 by native language and picture type. Reprinted from *Cognition*, 84, Anna Papafragou, Christine Massey, and Lila Gleitman, Shake, rattle, 'n' roll: The representation of motion in language and cognition, 189–219, copyright 2002, Figure 3, with permission from Elsevier.

the same pictures they'd seen before. Some of the pictures really were the same; others were similar to the earlier pictures but different in some crucial way. Some of the modified pictures changed the PATH of the original motion (e.g., a frog jumping out of a room instead of into it), while others changed the MANNER (e.g., a boy stumbling over a log instead of jumping over it). If Greek speakers pay more attention to PATH and English speakers to MANNER, then we might expect Greek speakers to do better at noticing pictures whose PATH had changed, while English speakers would do better at noticing pictures whose MANNER had changed.

As shown in Figure 11.3, subjects remembered pictures about equally well regardless of their native language. All subjects were better at correctly identifying pictures with no change (i.e., pictures they really had seen before) than rejecting pictures that had been changed, but there were no statistically significant differences between English and Greek speakers for either MANNER or PATH changes. These results, then, provide no evidence that English and Greek speakers think differently about motion events.

11.3.4 The direction of time: Boroditsky (2001)

Many languages use some of the same words to talk about space and time. In English, time is often imagined as an invisible horizontal line: the future is

ahead of us and the past is *behind* us; we move *forward* to new things and think *back* on what's happened before. Cross-linguistically, it's very common for words that originally had spatial meanings to acquire temporal meanings as well.

Mandarin has expressions that use the same front-back line to talk about time that English speakers do. But Mandarin speakers can also talk about time using a different imaginary axis – a vertical one. In this conception, the past is 'above' and the future is 'below'; examples are given in (3) and (4).

(3) Vertical metaphors: 'earlier' (Boroditsky 2001, Figure 2)

- a. māo shàng shù
'cats climb trees'
- b. shàng ge yuè
'last (or previous) month'

(4) Vertical metaphors: 'later' (Boroditsky 2001, Figure 2)

- a. tā xià le shān méi yǒu
'has she descended the mountain or not?'
- b. xià ge yuè
'next (or following) month'

Obviously, time isn't literally a horizontal or vertical line, but it's possible that spatial metaphors like these lead us to think about it that way. Boroditsky (2001) asked whether the specific metaphors a language uses could influence how speakers imagine time: since Mandarin has both horizontal and vertical metaphors, are Mandarin speakers more likely to think about time as a vertical line than English speakers, whose metaphors are almost exclusively horizontal?

To explore this question, Boroditsky used a psychological phenomenon known as *priming*. The basic idea is that people are often faster at some task if they've just seen or done something related; for example, if a subject's job is to recognize words, she will be faster at recognizing the word *doctor* if she's just seen a related word such as *nurse*. Boroditsky studied 26 native speakers of English and 20 native speakers of Mandarin; she asked them to answer true-false questions about the relative order of events (e.g., *March comes before April*). Two types of time words were used in these questions, either *before/after* (which some English speakers can use with a spatial meaning) or *earlier/later* (which are purely temporal). Interspersed with these time-related questions were spatial questions, in which subjects saw two objects and had to judge their relative location (some examples are given in Figure 11.4). Some of the spatial questions involved the horizontal axis, while others involved the vertical axis. Boroditsky predicted that English speakers would be primed to answer time-related questions after they had just

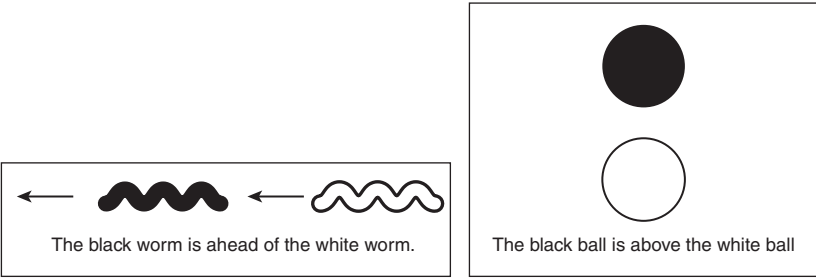


Figure 11.4 Examples of spatial questions used as primes. Reprinted from *Cognitive Psychology*, 43, Lera Boroditsky, Does language shape thought? Mandarin and English speakers' conceptions of time, 1–22, copyright 2001, Figure 3, with permission from Elsevier.

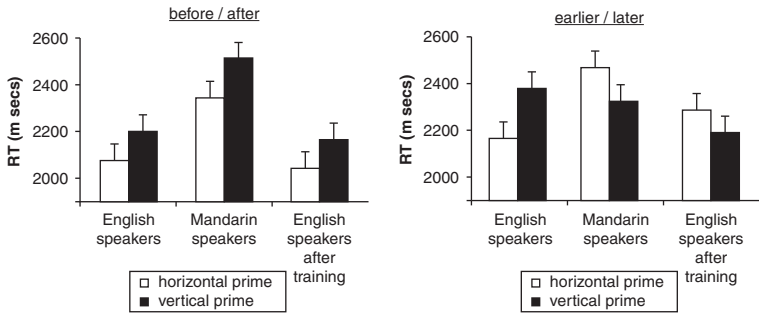


Figure 11.5 Reaction times to temporal questions by native language, type of spatial prime, and temporal language (*before/after* vs. *earlier/later*). Reprinted from *Cognitive Psychology*, 43, Lera Boroditsky, Does language shape thought? Mandarin and English speakers' conceptions of time, 1–22, copyright 2001, Figure 4, with permission from Elsevier.

answered horizontal spatial questions; in other words, thinking about horizontal space would help English speakers think about time because they imagine time that way. Mandarin speakers, on the other hand, might be primed by the vertical spatial questions, because their language makes a vertical metaphor available.

Figure 11.5 summarizes the results. For the time questions that used *before/after*, both English and Mandarin speakers answered faster if they had just seen a horizontal prime than if they had just seen a vertical prime; in other words, reasoning about horizontal space helped speakers of both languages reason about time. For the time questions that used *earlier/later*, the two groups were different: English speakers benefited from the horizontal

primes, as before, but Mandarin speakers benefited from the vertical primes. In other words, when they were answering questions about time that didn't use any spatial language at all, English speakers appeared to associate time with horizontal space, while Mandarin speakers associated time with vertical space.

Boroditsky conducted two other experiments to assess how much this effect really depends on language. In one experiment, she asked English speakers to complete an identical task – but before they did so, she trained them in what she called 'a new way to talk about time'. Subjects learned to use vertical spatial metaphors to describe the relative order of events, e.g., *Monday is above Tuesday*. As shown in Figure 11.5, English-speaking subjects who received this training performed like the Mandarin-speaking subjects of the original experiment: they answered *earlier/later* questions faster after vertical primes than after horizontal primes. Boroditsky concluded that even brief training in a different way of using language can cause people to think differently about time. In another experiment, Boroditsky asked 25 Mandarin-English bilinguals to complete part of the original experiment (with only the *earlier/later* questions included); she found that the priming effect of vertical spatial questions was bigger for subjects who had begun learning English later in life, and concluded that early exposure to English made the Mandarin speakers less likely to think about time as a vertical space.

Taken together, all of these experiments suggest that these metaphors do encourage speakers to think about time in particular ways; what we have, then, is support for a weak version of the Sapir-Whorf hypothesis. Interestingly, some of these same results also provide evidence *against* the strong version of the hypothesis: even though English speakers are apparently biased to think about time horizontally, they could be brought to think about time vertically after just a brief training period. In other words, a lifetime of using horizontal metaphors for time could be counteracted by just a little practice; English speakers were clearly not trapped by their habitual way of speaking.

We should also keep in mind that language isn't the only tool we have for representing time spatially. For English speakers, the idea of time as a horizontal line is reinforced in many other ways: the direction of writing, the traditional left-to-right layout of timelines, the arrangement of days within a week on a calendar, the horizontal control bars we use to navigate audio and video files, and so on. (We also have vertical representations of time, such as the arrangement of weeks on a calendar, but these are less common.) Confronted with all these ways of representing time, we might ask which way the causation goes: do horizontal metaphors cause English speakers to represent time horizontally in all these other domains too, or is language just one more consequence of a general cultural agreement that time is horizontal? Similarly, we might ask whether alternative forms of training could encourage English

speakers to think about time vertically. For example, suppose English speakers were trained to arrange the days of the week in a vertical layout – would this have had the same effect as the linguistic training?

Finally, it's important to note that although Boroditsky's study is well known and has been highly influential, it's also controversial. In particular, a number of other researchers (Chen 2007; January and Kako 2007; Tse and Altarriba 2008) have tried and failed to replicate the results, which raises the question of whether these apparent differences between English and Mandarin speakers are even real. Either there is no difference, and Boroditsky obtained a significant result by chance; or other labs used slightly different procedures that made the difference disappear – and if the latter is true, then the effect of language is apparently very fragile and unlikely to matter in real-world situations.

11.3.5 *Framing events: Loftus and Palmer (1974)*

Loftus and Palmer (1974) conducted a famous study on how language can be used to frame events in a way that affects how observers interpret those events. In one experiment, they showed several videos of traffic accidents to 45 students and then asked them about what they had seen. The crucial question was *About how fast were the cars going when they ___ each other?* Each subject saw a version of this question with one of five verbs: *hit*, *smashed*, *collided*, *bumped*, and *contacted*. As summarized in Table 11.1, Loftus and Palmer found that subjects' speed estimates depended in part on what verb they had seen in the question – their speed estimates were higher for more 'forceful' verbs, a significant difference at $p < .005$. Apparently, the way you ask someone about an event can affect the way that person remembers it.

In a second experiment, Loftus and Palmer asked 150 students to watch a single video of a traffic accident and answer questions about it. Some subjects were asked the same question as before, this time with one of two verbs (*hit* or *smashed*); other subjects weren't asked about the speed of the cars at all. One week later, the same subjects answered different questions about the accident; here, the crucial question was *Did you see any broken glass?* As shown in Table 11.2, Loftus and Palmer found that subjects were significantly ($p < .025$) more likely to answer *yes* if they had seen the verb *smashed* a week earlier than if they had seen the verb *hit* or hadn't been asked about speed at all – despite the fact that there was actually no broken glass in the video.

Did these different ways of phrasing the question actually affect how subjects remembered the accidents they had seen with their own eyes? Apparently so. Although we have seen that language doesn't put speakers in a straitjacket, it does seem possible for specific linguistic choices to 'nudge' people in one direction or another. In a similar study, Fausey and Boroditsky (2010) asked students about Janet Jackson and Justin Timberlake's 'wardrobe malfunction'

Table 11.1 *Mean speed estimates in response to questions with different verbs. Reprinted from Journal of Verbal Learning and Verbal Behavior, 13, Elizabeth F. Loftus and John C. Palmer, Reconstruction of automobile destruction: An example of the interaction between language and memory, 585–589, copyright 1974, Table 1, with permission from Elsevier.*

Verb	Mean speed estimate
Smashed	40.8
Collided	39.3
Bumped	38.1
Hit	34.0
Contacted	31.8

Table 11.2 *Subjects who responded yes and no to the question Did you see any broken glass? by verb condition. Reprinted from Journal of Verbal Learning and Verbal Behavior, 13, Elizabeth F. Loftus and John C. Palmer, Reconstruction of automobile destruction: An example of the interaction between language and memory, 585–589, copyright 1974, Table 2, with permission from Elsevier.*

Response	Verb condition		
	Smashed	Hit	Control
Yes	16	7	6
No	34	43	44

(see section 11.2.2); they found that subjects recommended larger monetary penalties when they read that Timberlake *tore the bodice* than when they read that *the bodice tore*, even when they had just watched a video of the incident.

However, as discussed above, we can understand these effects without concluding that language has some magical power over us. For obvious reasons, no politician will propose *An Act for Giving Tax Breaks to Corporations That Give Me Money*, even if that's what the bill does; similarly, cigarette manufacturers don't voluntarily cover their products with pictures of yellow teeth or tar-covered lungs. We are influenced by many things, and language is no exception.

11.3.6 General conclusions

Although some of the studies we have seen don't support any version of the Sapir-Whorf hypothesis, others suggest that a weak version may in fact be true: the way we perceive color, for example, may be subtly influenced by the language we speak. It's often hard to tell whether we're really seeing an influence of language on thought, as opposed to an influence of larger cultural patterns on both thought and language, but in at least some cases (e.g., grammatical gender in Spanish and German) it seems plausible that language is playing a causal role.

However, the evidence above also suggests that our understanding of the Sapir-Whorf hypothesis should be nuanced in at least two important ways. One is that most of the effects we've seen are very small – differences of less than a second in reaction times, or differences of a tenth of a point on a seven-point scale. These aren't the kinds of differences that would be obvious in everyday life; we can't point to a Spanish speaker, for example, and say, 'The only reason you think these apples are beautiful is because *manzana* is feminine in your language.'

The other important nuance is that, at least in some cases, the effect of language seems to be restricted to situations where the person is actually using language in some way, either consciously or not. English and Russian speakers, for example, behave the same when they're prevented from using language in a color-matching task. Language, it seems, hasn't altered speakers' color perception permanently; rather, it matters only for the kind of thought that is crucially connected to language – what Slobin (1991) calls 'thinking for speaking'.

The good news, then, is twofold. Language does have some very interesting effects on how people think, and working out exactly what those effects are will occupy researchers for decades to come. But the effects are subtle, and all the evidence suggests that we are *not* slaves to our language in the way Orwell feared. Newspeak isn't going to become a reality anytime soon.

11.4 Summary

- The *Sapir-Whorf hypothesis* is the idea that a person's language influences the way he or she thinks. The hypothesis comes in strong and weak versions.
- Many popular beliefs about language, especially in the realm of politics, assume a strong version of the Sapir-Whorf hypothesis: that language can make some thoughts unthinkable. However, there are many reasons to believe that the strong version of the hypothesis is not true.
- Several studies have found evidence for the weak version of the Sapir-Whorf hypothesis: language has a non-deterministic influence on some kinds of

thought. Some of these studies suggest that the effect is present only when people are actively using language to accomplish some task.

- It is often difficult in a given situation to determine whether language is really affecting how speakers think, or whether the way people think affects the language they use.
- There is good evidence that language can be used to ‘frame’ events and ideas, subtly influencing how people think about them; however, language is not the only tool that can be used in this way.

For further reflection

- (1) Choose a controversial political topic, and do an internet search for people explicitly comparing certain ways of talking about that topic to Newspeak. Analyze the discussions you find. Do commentators seem genuinely concerned that specific words or phrases will fool the public into accepting bad policies? Do you agree? Why or why not? To what extent do you think these terms can be used to frame public discussion in a way that influences people’s opinions? To what extent do you think people choose their words in order to show where they stand on the issue? To what extent could these words or phrases be characterized as ordinary lying?
- (2) Another example of linguistic diversity is the difference between languages that use *egocentric* directions (‘left’, ‘right’, ‘behind’, etc.) and those that exclusively use *cardinal* directions (‘north’, ‘south’, etc.). Read the extended discussion of this difference in Deutscher (2010a) or chapter 7 of Deutscher (2010b) and consider what it might imply about the Sapir-Whorf hypothesis. Do you find it plausible that speaking a language like Guugu Yimithirr might encourage a person to think differently about space? Do you think it’s possible that the language might reflect cultural convention? Think about your own experience with egocentric and cardinal directions: what kinds of situations make it easier or harder for you to keep track of directions such as ‘left’ or ‘north’?
- (3) Two more studies of the relationship between language and thought include Mazuka and Friedman (2000) (shape vs. material in English and Japanese) and Gennari et al. (2002) (motion in English and Spanish). Read and evaluate one of these papers. How did the authors go about testing thought independent of language, and to what extent do you think they succeeded? Are you convinced by the authors’ conclusions? In what ways, if any, might the two groups of speakers be different other than in their language?

Further reading

Two of Whorf's essays most often cited in connection with the Sapir-Whorf hypothesis are Whorf (1941) and Whorf (1956). Recent discussions of the psycholinguistic literature, written for a general audience, include Boroditsky (2010), Deutscher (2010b), and McWhorter (2014). The latter two books position themselves on opposite sides of the issue (Deutscher argues in favor of the hypothesis; McWhorter argues against it), but they actually come to the same conclusion: a weak version of the hypothesis is true, but the strong version is not. For an even more skeptical view, see Chapter 3 of Pinker (1994).

Orwell's famous essay on politics and language is Orwell (1968). Cameron (1995) is a thoughtful review of language use in the political sphere, particularly with regard to political correctness.

Bloom (1981) is a well-known early study of the Sapir-Whorf hypothesis, but with many methodological flaws; see Au (1984) for criticism. Boroditsky et al. (2003) reviews some of the literature, with a special focus on grammatical gender; however, several of the experiments described here seem to have never been peer-reviewed or published. Lucy and Gaskins (2001) and Imai and Gentner (1997) are experiments testing whether language affects how much a person pays attention to the shape of an object vs. the material it is made of. Fausey and Boroditsky (2011) is a study of whether language affects how well a person remembers the agents involved in an event.

Bibliography

- Aronson, Stanley M. Commentary: Refining the art of doublespeak. *The Providence Journal*, pages C–05, March 26 2007.
- Au, Terry Kit-Fong. Counterfactuals: In reply to Alfred Bloom. *Cognition*, 17(3):289–302, 1984.
- Berlin, Brent, and Paul Kay. *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley, CA, 1969.
- Bloom, Alfred. *The Linguistic Shaping of Thought: A Study in the Impact of Language on Thinking in China and the West*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1981.
- Boroditsky, Lera. Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, 43(1):1–22, 2001.
- Boroditsky, Lera. Lost in translation. *The Wall Street Journal*, July 23 2010. Available at <http://www.wsj.com/articles/SB10001424052748703467304575383131592767868>.
- Boroditsky, Lera, Lauren A. Schmidt, and Webb Phillips. Sex, syntax, and semantics. In Dedre Gentner and Susan Goldin-Meadow, editors, *Language in Mind: Advances in the Study of Language and Thought*, chapter 4, pages 61–79. The MIT Press, Cambridge, MA, 2003.

- Broomstreet, Henry. Letter to the editor: Abortion doublespeak. *San Bernardino Sun*, March 13 2006.
- Cameron, Deborah. *Verbal Hygiene*. Routledge, London, 1995.
- Chen, Jenn-Yeu. Do Chinese and English speakers think about time differently? Failure of replicating Boroditsky (2001). *Cognition*, 104(2):427–436, 2007.
- Deutscher, Guy. *Through the Language Glass: Why the World Looks Different in Other Languages*. Metropolitan Books, New York, NY, 2010a.
- Deutscher, Guy. Does your language shape how you think? *The New York Times*, August 29 2010b. Available at <http://www.nytimes.com/2010/08/29/magazine/29language-t.html?fta=y>.
- Fausey, Caitlin M., and Lera Boroditsky. Subtle linguistic cues influence perceived blame and financial liability. *Psychonomic Bulletin & Review*, 17(5):644–650, 2010.
- Fausey, Caitlin M., and Lera Boroditsky. Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin & Review*, 18(1):150–157, 2011.
- Gennari, Silvia P., Steven A. Sloman, Barbara C. Malt, and W. Tecumseh Fitch. Motion events in language and cognition. *Cognition*, 83(1):49–79, 2002.
- Heider, Eleanor Rosch. Universals in color naming and memory. *Journal of Experimental Psychology*, 93(1):10–20, 1972.
- Heider, Eleanor Rosch, and Donald C. Olivier. The structure of the color space in naming and memory for two languages. *Cognitive Psychology*, 3(2):337–354, 1972.
- Hentoff, Nat. Euthanasia: Another warning from the Surgeon General. *The Washington Post*, page A21, April 11 1987.
- Imai, Mutsumi, and Dedre Gentner. A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, 62(2):169–200, 1997.
- January, David, and Edward Kako. Re-evaluating evidence for linguistic relativity: Reply to Boroditsky (2001). *Cognition*, 104(2):417–426, 2007.
- Kershaw, Ian. How effective was Nazi propaganda? In David Welch, editor, *Nazi Propaganda: The Power and the Limitations*, chapter 10, pages 180–205. Croom Helm, London, 1983.
- Konishi, Toshi. The semantics of grammatical gender: A cross-cultural study. *Journal of Psycholinguistic Research*, 22(5):519–534, 1993.
- Liberman, Mark. 'No word for X' archive. Language Log, January 28 2009. Available at <http://languagelog.ldc.upenn.edu/nll/?p=1081>.
- Loftus, Elizabeth F., and John C. Palmer. Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 13(5):585–589, 1974.
- Lucy, John A. The scope of linguistic relativity: An analysis and review of empirical research. In John J. Gumperz and Stephen C. Levinson, editors, *Rethinking Linguistic Relativity*, number 17 in Studies in the Social and Cultural Foundations of Language, chapter 2, pages 37–69. Cambridge University Press, Cambridge, 1996.
- Lucy, John A., and Suzanne Gaskins. Grammatical categories and the development of classification preferences: A comparative approach. In Melissa Bowerman and Stephen C. Levinson, editors, *Language Acquisition and Conceptual Development*, number 3 in Language, Culture and Cognition, chapter 9, pages 257–283. Cambridge University Press, Cambridge, 2001.

- Mazuka, Reiko, and Ronald S. Friedman. Linguistic relativity in Japanese and English: Is language the primary determinant in object classification? *Journal of East Asian Linguistics*, 9(4):353–377, 2000.
- McWhorter, John H. *The Language Hoax: Why the World Looks the Same in Any Language*. Oxford University Press, Oxford, 2014.
- Orwell, George. 1984. Harcourt Brace Jovanovich, New York, NY, 1949.
- Orwell, George. Politics and the English language. In Sonia Orwell and Ian Angus, editors, *The Collected Essays, Journalism and Letters of George Orwell*, volume 4: In Front of Your Nose, 1945–1950, pages 127–140. Penguin Books, London, 1968.
- Papafragou, Anna, Christine Massey, and Lila Gleitman. Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition*, 84(2):189–219, 2002.
- Pinker, Steven. *The Language Instinct: How the Mind Creates Language*. William Morrow and Company, New York, NY, 1994.
- Poser, Bill. No word for thank you. *Language Log*, May 6 2006. Available at <http://itre.cis.upenn.edu/~myl/language-log/archives/003120.html>.
- Pullum, Geoffrey K. *The Great Eskimo Vocabulary Hoax and Other Irreverent Essays on the Study of Language*, chapter 19: The Great Eskimo Vocabulary Hoax, pages 159–171. The University of Chicago Press, Chicago, IL, 1991.
- Qiang, Xiao. The grass-mud horse lexicon. Available at http://chinadigitaltimes.net/space/The_Grass-Mud_Horse_Lexicon.
- Roberson, Debi, Ian Davies, and Jules Davidoff. Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, 129(3):369–398, 2000.
- Slobin, Dan I. Learning to think for speaking: Native language, cognition, and rhetorical style. *Pragmatics*, 1(1):7–25, 1991.
- Talmy, Leonard. Lexicalization patterns: Semantic structure in lexical forms. In Timothy Shopen, editor, *Language Typology and Syntactic Description*, volume 3: Grammatical Categories and the Lexicon, chapter 2, pages 57–149. Cambridge University Press, Cambridge, 1st edition, 1985.
- Tse, Chi-Shing, and Jeanette Altarriba. Evidence against linguistic relativity in Chinese and English: A case study of spatial and temporal metaphors. *Journal of Cognition and Culture*, 8(3/4):335–357, 2008.
- Whorf, B. L. The relation of habitual thought and behavior to language. In Leslie Spier, A. Irving Hallowell, and Stanley S. Newman, editors, *Language, Culture, and Personality: Essays in Memory of Edward Sapir*, pages 75–93. Sapir Memorial Publication Fund, Menasha, WI, 1941.
- Whorf, Benjamin Lee. Science and linguistics. In John B. Carroll, editor, *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, pages 207–219. The MIT Press, Cambridge, MA, 1956.
- Winawer, Jonathan, Nathan Witthoft, Michael C. Frank, Lisa Wu, Alex R. Wade, and Lera Boroditsky. Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, 104(19):7780–7785, 2007. Available at <http://www.pnas.org/content/104/19/7780.full>.