Casasanto, D. (2025). Sapir-Whorf Hypothesis. In Sinha, C. & Xu, W. (Eds.) *The Cambridge Encyclopedia of Cognitive Linguistics.* Cambridge, Cambridge University Press.

Sapir-Whorf Hypothesis

Daniel Casasanto

Daniel Casasanto
Cornell University
Martha Van Rensselaer Hall
Ithaca, NY 14850 USA
casasanto@alum.mit.edu

Abstact

According to the Sapir-Whorf Hypothesis, language affects the way people think, causing speakers of different languages to think differently. Formulated in the early 20th century, this hypothesis has remained controversial across linguistics, psychology, philosophy, and anthropology. Although the Sapir-Whorf Hypothesis was widely thought to be discredited at the end of the 20th century, new experimental evidence in support of the hypothesis has accumulated since the turn of the 21st century. This article briefly reviews the history of this hypothesis. It then distinguishes two formulations of the hypothesis known as *linguistic relativity* (i.e., the proposal that habits of language influence habits of thinking) and *linguistic determinism* (i.e., the proposal that features of language rigidly determine the thoughts people can and cannot have) and summarizes the current evidence for each of these formulations. Finally, psychological mechanisms by which language shapes thinking are discussed, as are potential implications of linguistic relativity and linguistic determinism for understanding the mind and the world.

Keywords: Linguistic relativity; Linguistic determinism; Cognitive universals; Cognitive diversity; Linguistic typology

Sapir-Whorf Hypothesis

1. *Introduction and historical perspectives*

According to the Sapir-Whorf Hypothesis, language shapes how people construct reality. This shaping process occurs no matter what language people speak, but since each language highlights different aspects of the world, people who speak different languages come to think differently: construing the world in ways that are consistent with the words and grammatical structures in their languages. Today, this hypothesis is associated most often with Benjamin Lee Whorf (1897-1941) and secondarily with his mentor, the anthropological linguist Edward Sapir (1884-1939). Earlier versions of this idea were formulated by the philosopher Wilhelm von Humboldt (1767-1835) and the psychologist Wilhelm Wundt (1832-1920).

Neither Sapir nor Whorf ever stated any formal hypothesis regarding relationships between language and thought, but their views were nevertheless articulated in their writings which, for Whorf, were mostly published for the first time posthumously in 1956. Whorf posited that much of the content of the human mind "is not linguistic but yet shows the shaping influence of language." (1939 / 1956, pg. 147). Whorf points to three things that led him to develop this view. First, he was influenced by Sapir's arguments that "human beings do not live in the objective world alone…but are very much at the mercy of the particular language which has become the medium of expression for the society. […] The 'real world' is to a large extent built up of the language habits of the group" (Sapir, 1929, pg. 209).

Second, Whorf drew on his professional experiences, not as a linguist or any other kind of academic researcher, but rather as an engineer who investigated accident claims for the Hartford Fire Insurance Company. In this role, Whorf observed many human errors that he attributed to the language that people used. For example, he investigated an explosion caused by someone tossing

a lit cigarette butt into an "empty" gasoline drum. Whorf argued that this error resulted from English speakers' natural tendency describe a drum that had been drained of its liquid contents as "empty" – even though it was then full of combustible vapor. On this view, labeling gasoline drums as either "liquidful" or "vaporful" could prevent such accidents.

Finally, Whorf conducted influential research in anthropological linguistics, analyzing relationships between language and culture, and extrapolating to links between language and the mind. Although he held a brief appointment as a lecturer in Yale University's Anthropology department, Whorf conducted most of this research as an amateur with no academic affiliation, nor any relevant advanced degrees (Whorf held a Bachelor of Science degree in chemical engineering from the Massachusetts Institute of Technology). Many of Whorf's conclusions are still being debated nearly a century later. He raised the question: "Are our own concepts of 'time', 'space', and 'matter' given in substantially the same form by experience to all men, or are they in part conditioned by the structure of particular languages?" (1939 / 1956, pg. 138). In some of his most controversial claims Whorf argued that, compared to "Standard Average European" (SAE) languages like English, Native American languages confer on their speakers different conceptions of time, and of matter, and of relationships between time and matter (i.e., between events and objects).

In English, for example, the grammatical categories *noun* and *verb* correspond to conceptual categories of *objects* and *events.* It is tempting to assume that the distinction between these categories, in language and thought, reflects a basic, universal, ontological distinction in the world. Whorf's analyses suggest otherwise. Whorf argued that it is not possible to draw this ontological distinction on the basis of nature, and that people's beliefs about whether an entity is an object or an event depend on whether it is named by a noun or a verb in their language. He

notes that entities like "[a] wave, [a] flame, [or a] meteor" are all named by nouns in English, and are presumably construed as objects, whereas these entities are named by verbs in Hopi, and are (just as sensibly) construed as events. In Nootka, Whorf notes, all entities are named by verbs, even entities that to an English speaker may seem to be parade cases of stable concrete objects. The Nootka equivalent of *a house* would be "a house occurs" or "it houses" (Whorf, 1941/1956, pg. 215-216); a house is not an object, *being a house* is a protracted event.

Whorf noted multiple differences between the encoding of time in Hopi and English that suggest differences in how time is conceptualized. For example, in English the duration of an event is typically described as a "length of time," but in Hopi, "our 'length of time' is not regarded as a length but as a relation between two events in lateness" (Whorf, 1939/1956, pg. 140). Whorf elaborated on the tendency to use spatial metaphors in English, and by inference in English speaker's thoughts, which he reported to be absent from Hopi:

> "[W]e can hardly refer to the simplest nonspatial situation without constant resort to physical metaphors. I 'grasp' the 'thread' of another's arguments, but if its 'level' is 'over my head' my attention may 'wander' and 'lose touch' with the 'drift' of it, so that when he 'comes' to his 'point' we differ 'widely', our 'views' being indeed so 'far apart' that the 'things' he says 'appear' 'much' too arbitrary, or even 'a lot' of nonsense! The absence of such metaphor from Hopi speech is striking. The use of space terms when there is no space involved is NOT THERE (Whorf, 1939/1956, pg. 146, emphasis in the original text).

Rather than using spatial words like ‘long’, Hopi indicates duration and other temporal concepts via their grammar, using morphological devices such as verbal aspect and tensors which denote time but not space. Accordingly, Whorf posited that Hopi speakers did not conceptualize time analogically, as imaginary space, but rather conceptualized it directly, *qua* time.

On the basis of such differences between languages, and by inference in the minds of their speakers, Whorf advanced what he called a “principle of linguistic relativity” (Whorf, 1940/1956, pg. 214): People’s habits of conceptualizing the world vary relative to their habits of describing it in language.

2. *Critical reception to Whorf’s claims*

Whorf’s writings were received with a level of hostility that is rare in academic discourse, which was sustained for decades (e.g., see Pullum, 1991). One reason for this response may be that Whorf’s claims ran contrary to two intellectual trends in the mid 20th century. First, in the wake of the anthropologist Margaret Mead’s controversial book *Coming of Age in Samoa* (1928), social scientists became highly sensitized to Western scholarship that could be seen as exoticizing non-industrialized cultures. Second, a movement in psychology and linguistics began, called the ‘cognitive revolution’, initiated in part by writings of the linguist Noam Chomsky (Chomsky, 1957; 1965). The previous movement in psychology, known as Behaviorism (Watson, 1928), prioritized the role of an individual’s experience in shaping the mind. By contrast, the cognitive revolutionaries’ theories prioritized aspects of mind and language that were posited to be innate and universal. Whorf’s claims have been interpreted as inconsistent with a ‘Universalist’ view of mind and language.

In an influential critique, the psychologist Steven Pinker devoted a chapter of his 1994 book *The Language Instinct* to criticizing – and ostensibly debunking – Whorf's claims. This chapter summarized theoretical and empirical arguments other scholars had made against Whorf since the 1950s. In the three decades following the publication of *The Language Instinct*, many of these arguments were rebutted.

Pinker's (1994) primary objection to Whorf was theoretical. According to Pinker (1994), Whorf believed that "thought is the same thing as language" (pg. 57). After providing compelling evidence that at least some human thinking is nonlinguistic, Pinker (1994) concluded that the Sapir-Whorf hypothesis was "wrong, all wrong" (pg. 57). Yet, as noted by the psychologist Daniel Casasanto (2008), Whorf never argued that "thought is the same thing as language." On the contrary, Whorf believed that much of human thinking is "nonlinguistic" (1939/1956, pg. 147; see above), but still susceptible to the shaping influence of language. Therefore, Pinker's (1994) primary theoretical criticism of Whorf was invalid.

Pinker's empirical critique pointed to problems with Whorf's linguistic analyses, in particular his analyses of Hopi, citing contrasting analyses of Hopi by the anthropological linguist Ekkehart Malotki (1983). Malotki is often credited with debunking the claim that the Hopi had 'no conception of time'. Yet, the most extreme claims that Malotki (1983) decried, concerning "the 'timelessness' of Hopi life and language" (pg. 3), were not written by Whorf himself but rather by a commentator on Whorf, John Carroll (1956). Malotki (1983) accurately described Whorf's actual claim, which was far less radical: that the Hopi have "a different view of time" from SAE speakers (pg. 2). Opposing one of Whorf's most surprising claims, Malotki argued that Hopi did, in fact, contain spatial metaphors for non-spatial concepts. In considering these conflicting claims, it is notable that an important change occurred during the decades between Whorf's analyses and

Malotki's: The number of monolingual Hopi speakers declined dramatically, and the influence of colonizing languages such as English and Spanish increased (Hill, 1997). Therefore, Malotki's analysis of spatial metaphors in Hopi could have included loan words and constructions from SAE languages that were not used by previous generations of Hopi speakers. Malotki's Hopi analyses, and their applicability to Whorf's claims, have been challenged more broadly by the linguistic anthropologist John Lucy (1992).

Echoing earlier scholars (Brown & Lenneberg, 1954; Pullum 1991), Pinker (1994) argued that it is not possible to show effects of language on thought based on linguistic analyses, alone. Whorf's analyses, therefore, can only serve as a source of *hypotheses* about how language might affect thinking. To avoid circular reasoning, hypotheses based on language must be tested with language-independent psychological data – which Whorf and his contemporaries lacked. Starting in the mid-20th century, a small number of psychological studies tested Whorfian questions, but Pinker (1994) and others dismissed these studies on two grounds. First, some studies' conclusions were invalid because of methodological flaws, such as clumsy translations that caused members of one language group to perform tasks more poorly than another. Second, some studies showed valid cross-linguistic differences but only tested "weak" and "banal" versions of the Sapir-Whorf hypothesis. In the late 20th and early 21st centuries, both of these concerns were addressed by a growing body of psychological studies.

3. *Psychological evidence for the Sapir-Whorf hypothesis*

3.1 Turning Whorf's ideas into experiments.

A first step in testing the Sapir-Whorf hypothesis experimentally is to specify exactly what theoretical claim is being tested. Two versions of the hypothesis were framed by 20th century

scholars: *linguistic relativity* and *linguistic determinism.* According to the linguistic relativity hypothesis, language influences how people think. This formulation is in line with Whorf's own "principle of linguistic relativity" (1940/1956): Habits of language use give rise to habits of thinking. According to the linguistic determinism hypothesis, language rigidly determines the thoughts people can have. Both linguistic relativity and linguistic determinism entail that language plays a causal role in thinking; the hypotheses differ in the extent to which a particular kind of thinking should be possible without language – whether language is *necessary* for some categories of thought. The majority of experiments to date have tested linguistic relativity (for an exception see section 3.4).

A second step in testing the Sapir-Whorf hypothesis is to specify what constitutes 'language' and 'thought' in the context of a given experiment: *What aspect of language is hypothesized to affect what aspect of thinking, and by what psychological mechanism?* Answering these questions allows an experimenter to 'operationalize' the Sapir-Whorf hypothesis: to turn an abstract idea into a set of concrete experimental procedures and measures.

3.2 Linguistic relativity of color

Roger Brown and Eric Lenneberg (1954) first operationalized the linguistic relativity hypothesis in the domain of color. They noted that colors are not labeled the same way across different languages. The Iakuti language, for example, has a single word for the colors that English speakers would call green and blue. Within a given language, colors differ in what Brown and Lenneberg called "codability": Highly codable colors are given the same name consistently by different speakers, whereas less codable colors are given different names across speakers. Brown and Lenneberg posited that: (i.) highly codable colors get labeled by speakers more frequently than

low-codability colors; (ii.) frequently labeling a color causes the color to be more psychologically "available" in speakers' minds; (iii.) therefore, highly codable colors should be easier to remember than low-codability colors (controlling for the perceptual discriminability of the colors). Consistent with this prediction, native English-speaking participants were more successful at recognizing highly codable color chips that they had been shown previously, picking out the colors they had studied from a large matrix of color chips. The effect of codability on memory was greater when the delay between studying the colors and recognition testing was increased, presumably because verbal labels become more important as the perceptual memory of the colors fades.

Brown and Lenneberg's (1954) within-language study in English speakers was corroborated by a cross-linguistic adaptation of their codability study. Color memory was compared between English speakers and speakers of Zuni, which has a single term for the English color categories orange and yellow. Monolingual Zuni speakers frequently confused orange and yellow stimuli, which were never confused by English speakers. Zuni speakers who were bilingual in English showed intermediate recognition memory (Lenneberg & Roberts, 1953). Together, these studies provided the first clear validation of the linguistic relativity hypothesis. Brown and Lenneberg concluded that "there may be general laws relating [linguistic] codability to [nonlinguistic] cognitive processes" (pg. 462).

Critical attitudes toward the Sapir-Whorf hypothesis were changed very little by Brown and Lenneberg's (1954) findings. Pinker (1994) voiced one of the common responses to their study: "[T]hat words can have some effect on memory […]is hardly surprising." To address the objection that Brown and Lenneberg's (1954) codability effects were 'just on memory', not on color perception per se, a team led by psychologists Jonathan Winawer and Lera Boroditsky ran a speeded color discrimination study. Participants saw three blue squares arranged in a triangle with

one square on top and the other two squares below it. One of the lower squares was the same shade of blue as the top square, and the other bottom square was a slightly different shade of blue. Participants' task was to indicate which of the lower squares was identical to the top square. Although all of the squares would be coded as "blue" by an English speaker, the shades of blue spanned two distinct color categories in Russian: *goluboy* (light blue) and *siniy* (dark blue). Russian speakers, but not English speakers, were faster to discriminate the colors when they crossed the goluboy/siniy border. To determine the timecourse on which language was affecting color discrimination, Winawer et al. (2011) ran a version of their color matching task with concurrent verbal interference: Participants silently rehearsed a string of digits while performing the task, to block their ability to code the colors in words. Verbal interference eliminated the benefit of crossing the goluboy/siniy border for Russian speakers, indicating that participants in the no-interference condition were activating color terms silently when matching colors.

Together, these and related studies show that the repertoire of color terms in one's native language can affect not only how well people remember colors but also how they make basic perceptual discriminations among colors – a process long believed to rely on a 'color module' in the human brain that was impervious to influences from higher-level cognitive systems such as language (Pylyshyn, 1999).

3.3 Linguistic relativity of time

Although no psychological tests have investigated Whorf's claims about time concepts in the Hopi, linguistic relativity effects have been reported in speakers of several other languages. In one seminal study, Lera Boroditsky (2001) compared mental representations of temporal sequences between speakers of English and Mandarin. In both English and Mandarin, spatial

metaphors suggest that time flows horizontally (e.g., the future is *ahead* and the past *behind* the speaker). Unlike English, Mandarin also uses vertical metaphors for time: Earlier times are 'up' (*shàng*) and later times are 'down' (*xià*). In a series of studies, native English and native Mandarin speakers were psychologically 'primed' to think about either horizontal or vertical spatial scenarios before responding to temporal sentences (e.g., March comes earlier than April). To ensure that any differences in performance between the language groups were not the result of inexact translations, all stimuli were presented in English, and all Mandarin-speaking participants were fluent speakers of English as a second language. Results showed that English speakers benefitted more from the horizontal primes whereas Mandarin speakers benefitted more from the vertical primes, responding faster to the 'earlier/later' sentences after judging a vertical spatial scene. Mandarin speakers who started learning English at an earlier age showed less benefit from vertical primes than Mandarin speakers who learned English later in childhood, suggesting that early exposure to English caused participants to process temporal sentences in a less 'Mandarin-like' way.

These studies comparing native English and Mandarin speakers show correlations between features of language (i.e., the presence of vertical spatial metaphors) and ways of thinking (i.e., making judgments about events). But these studies, like all cross-linguistic or cross-cultural studies, are necessarily 'quasi-experimental': Participants are not randomly assigned to 'treatments' (i.e., being an English speaker or a Mandarin speaker). Quasi-experiments are necessary to characterize patterns of thinking in the natural world, but they cannot demonstrate causal effects of one experimental variable on another. To test for a causal effect of language on thought, Boroditsky (2001) conducted a 'true experiment' in which English speakers were trained to talk about time using Mandarin-like vertical metaphors (e.g., Nixon was president *above*

Clinton). After training, English speakers performed the experiment described above, and showed a vertical prime advantage for earlier/later sentences that was statistically indistinguishable from the native Mandarin speakers'. This experiment demonstrated a causal effect of using vertical spatial metaphors on people's temporal judgments. Moreover, the true experiment ruled out potential alternative explanations for the cross-linguistic differences found in the quasi-experiments. In principle, the Mandarin-English differences observed could have been caused by some aspect of the Mandarin speakers' culture (e.g., exposure to a vertical writing system), not by their language. In the training study, however, all aspects of culture were held constant (including the writing system); the only thing that changed in trained participants' experience was their use of vertical spatial metaphors for time.

Boroditsky's (2001) study was controversial, and two papers reported failing to replicate its findings (Chen, 2007; January & Kako, 2009). Subsequently, however, the finding that Mandarin speakers make greater use of the vertical axis for temporal sequences than English speakers was validated, both by Boroditsky and colleagues (2011) and by other researchers (Miles et al., 2011).

Another study by the cognitive linguists Rafael Nuñez and Eve Sweetser (2006) documented a surprising pattern of language and thought in the Aymara people of the Andean highlands. Across numerous languages, spatial metaphors suggest that the future is ahead and the past is behind (Alverson, 1994). This pattern was long thought to be a human universal, which was assumed to be based on immutable aspects of our anatomy and physiology (H. Clark, 1973; Lakoff & Johnson, 1999): Given the frontward orientation of our feet and sensory organs, people typically more *forward* along spatial paths. As a result of traveling forward, objects that we will encounter in the future lie ahead of us in space, and objects that we passed previously lie behind us. Yet,

despite these bodily universals, in Aymara the spatial word for 'front' (*nayra*) also means *past* and the spatial word for 'back' (*qhipa*) also means *future.* A clue about the motivation for this 'reversed' mapping of past and future comes from the fact that *nayra* means 'eye' and 'sight' as well as 'front'. Aymara informants explained that the past is in front because only the past can be 'seen' (i.e., known). Beyond documenting this pattern in language, Nuñez and Sweetser observed Aymara speakers gesturing forward when talking about the past and backward when talking about the future, suggesting that they conceptualize time the same way they talk about it.

A possible critique of these studies of metaphorical language and thought in Mandarin and Aymara concerns the role of language in the experimental tasks. In Boroditsky's (2001) studies participants were judging temporal sentences, and in Nuñez and Sweetser's (2006) participants were talking about events while gesturing. Therefore, participants were engaged in a process the psycholinguist Dan Slobin (1996) called "thinking for speaking." Slobin posited that when people package thoughts into words they engage in "a special kind of thinking," which may differ from the thinking they do when they are not using language. In line with Slobin's proposal, skeptics such as the linguist Eve Clark proposed that, despite emerging data showing thinking-for-speaking effects, the role of language in thought should be limited:

> "[W]e should find that in tasks that require reference to representations in memory that don't make use of any linguistic expression, people who speak different languages will respond in similar, or even identical, ways. That is, representations for nonlinguistic purposes may differ very little across cultures or languages (E. Clark, 2003, pg. 22).

Studies led by Daniel Casasanto and Lera Boroditsky (2008, 2025) demonstrated that language affects people's temporal thinking even when they are not 'thinking for speaking'. In one set of experiments, English and Greek speakers were given non-linguistic tests of their ability to estimate duration. As noted by Whorf (1939/1956), English tends to express duration in terms of spatial length (e.g., a *long* time). By contrast, Greek tends to express duration in terms of volume or amount (e.g., *a lot* of time (*poli* ora)). Participants were asked to reproduce the durations of stimuli they saw on a computer screen (i.e., lines gradually extending across the screen or containers gradually filling up) while ignoring the spatial extent of the lines or the fullness of the containers. English speakers had difficulty screening out irrelevant information about spatial length when estimating duration: Lines that traveled a longer distance were mistakenly judged to take a longer time than lines that traveled a shorter distance. But their time estimates were relatively unaffected by irrelevant volume information. Greek speakers showed the opposite pattern: They had more difficulty screening out irrelevant volume information, so fuller containers were erroneously judged to remain on the screen for more time than emptier containers. The pattern of length and volume interference in these nonlinguistic psychophysical tasks reflected the relative prevalence of length and volume metaphors for duration in English and Greek.

Although it is possible that participants in these studies were subvocally labeling the stimuli (e.g., silently coding the durations as 'long' or 'short'), the design of the stimuli ensured that subvocal labeling could not account for the results (i.e., spatial lengths and temporal durations varied orthogonally). To confirm that silent labeling was not driving the observed effects, the experimenters ran a version of the duration estimation task with concurrent verbal interference. Participants rehearsed a string of digits while perceiving the stimuli and reproducing their durations. Verbal interference did not affect the results: English speakers incorporated spatial

length information into their time estimates to the same extent even when they were prevented from coding the stimuli in language (cf., Winawer et al., 2011).

The experimenters proposed that repeatedly using either length metaphors or volume metaphors for time in language strengthens the association in long-term memory between non-linguistic representations of time and length, or between time and volume, while weakening the other association (Song et al., 2000). In this way, habitual use of linguistic metaphors affected people's *subsequent* thinking, and using language in the moment was not necessary to produce the different patterns of temporal thinking found in English and Greek speakers.

To test whether using volume metaphors in language can change the way people think about duration, Casasanto and colleagues trained English speakers to use Greek-like metaphors for time and then gave the trained participants one of the non-linguistic duration estimation tasks described above (i.e., the filling containers task). After about 20 minutes of intensive practice using Greek-like volume metaphors, the effect of irrelevant volume information on English speakers' non-linguistic duration estimates was statistically indistinguishable from the effect found in native Greek speakers. This laboratory training effect was complemented by a natural 'training effect': Greek speakers living in the US, and speaking English regularly, showed an effect of length on time estimation that was statistically indistinguishable from the effect found in US English speakers. It is notable that the capacity to reproduce brief approximate durations tested in these studies develops during infancy and is shared with non-human animals. Using one kind of spatial metaphor or another to talk about duration can shape even our most basic representations of time.

These studies are representative of the body of research produced in the first decades of the 21st century showing that the linguistic encoding of time varies across languages, and that temporal thinking varies accordingly. Effects of people's previous language use on their time

representations can be found even when participants are prevented from using language to perform the task. Laboratory training experiments show that the effects of language can be distinguished from the potential effects of non-linguistic cultural artifacts and practices, and that language can play a causal role in shaping mental representations of time.

3.3 From linguistic relativity to linguistic determinism

Color and time are only two among the growing number of conceptual domains in which 'Whorfian' effects have been shown. Although space here precludes reviewing all of the relevant literatures, language has been found to influence thinking in domains such as *space* (Levinson, 2003; Majid et al., 2004), *music* (Dolscheid et al., 2013; Patel & Daniele, 2003); *motion events* (Papafragou et al., 2008), *geometry* (Hermer & Spelke, 1996), *theory of mind* (Lohmann & Tomasello, 2003), *causation* (Fausey & Boroditsky, 2011), and *objects* (Boroditsky et al., 2003; Cubelli et al., 2011). Findings in these domains demonstrate linguistic relativity. To date, however, there is only one clear, well-supported example of linguistic determinism, in the domain of *number.*

3.4 Linguistic determinism of number

Children are not born with the capacity to represent 'large' exact quantities, meaning quantities greater than 3, nor do they develop this capacity through universal aspects of physical and social experience. They develop the capacity to mentally represent 'exactly 4,' or 'exactly 17,' studies suggest, only if they are exposed to a list of counting numbers in their language (e.g., 1, 2, 3, 4, 5...).

It is easy to take the existence of a verbal count list for granted since it is found in the languages used by all modern, industrialized cultures. Yet, counting systems like ours are recent and rare in human history, and are still unknown to people in many cultures (Dehaene, 1999). The impact of counting words on our minds and world was not understood before the early 21st century.

3.4.1 Pre-linguistic number systems

Humans begin life with two primitive numerical systems, which are shared with other non-human animals: the parallel individuation system and the approximate number system (Carey, 2004; Spelke, 2000). The parallel individuation system allows infants to recognize and differentiate one, two, or three objects, but no more. Even four objects exceed the range of this system.

The approximate number system enables infants to evaluate and compare larger quantities, but only in terms of relative amounts, not exact numbers. Its accuracy depends on the ratio between the quantities compared. For example, infants under six months can differentiate between quantities with a 1:2 ratio (e.g., 8 vs. 16) but not smaller ratios (e.g., 8 vs. 10). Adults can distinguish somewhat closer ratios than infants, but the approximate number system never reaches the precision needed to represent exact quantities, or to reliably distinguish between collections of 9 versus 10 objects.

These two primitive number systems, separately or in combination, lack the range and representational precision of symbolic counting, which must be acquired through language.

3.4.2 How number words lead to number concepts

Psychologist Susan Carey (2004) proposed a lengthy developmental process by which children use number words to create a new set of concepts, which was not present earlier in their development and which only emerges if children are taught to use number words. Children first learn to say the numbers as a word game (typically the numbers one through ten in Western cultures), much like they learn nonsensical nursery rhymes (e.g. eeney, meeny, miney, mo): They can recite the numbers in order, but they do not understand them. By having the spoken numbers matched with sets of fingers or other objects, children start to understand that each number word refers to a precise quantity. First, they learn that "one" refers to one object. Weeks or months later they learn that "two" refers to two objects. At this stage of being a 'two-knower,' children asked for "two marbles" can deliver the correct number, but when asked for "three marbles" or "four marbles" they will respond with some small collection of marbles greater than two. Eventually, after a period as 'three-knowers', children induce how counting works.

Carey (2004) refers to the process of using number words to create new number concepts as "bootstrapping", to acknowledge that children's number learning is not a quantitative change (e.g., making approximate number representations more precise) but rather a qualitative change in the nature of their concepts. Number concepts that correspond to productive use of counting words have the properties of 'cardinality' (i.e., exact numerical value) and 'ordinality' (i.e., they stand in fixed succession with respect to one another) – both properties are absent from children's primitive number systems. These properties allow number to be a productive system that makes infinite use of finite means, and to form the basis for countless mathematical operations.

3.4.3 Evidence for the linguistic bootstrapping of number concepts

Psychologist Peter Gordon (2004) provided the first evidence supporting the linguistic bootstrapping of number concepts. The Pirahã people of the Amazon speak a language that lacks words for exact numbers. Pirahã speakers use quantifiers like *hói* (about one), *hoí* (about two), and *baágiso* (many), but like other languages spoken by non-industrialized people, Pirahã has no list of counting words. When Gordon asked Pirahã adults to perform simple enumeration tasks, such as matching the number of nuts dropped into a can, the Pirahã participants frequently made mistakes that are vanishingly rare for Western adults. When the task required distinguishing four or more items, the majority of the Pirahã participants' responses were incorrect.

Gordon concluded that: (i.) the Pirahã are unable to represent exact numbers greater than three, and (ii.) this limitation results from the absence of number words in their language. These claims constitute a strong form of linguistic determinism: Counting words enable people to acquire number concepts, and the lack of these words causes people to lack these concepts.

Gordon's study faced criticisms concerning the data themselves and the inferences they could support. One feature of the data was particularly worrisome. Some of Gordon's tasks, like the nuts-in-a-can task, required exact enumeration. Other tasks, however, could be completed without counting, using a non-numerical 'one-to-one correspondence' strategy (e.g., placing one object on the table for each object that the experimenter placed on the table). The Pirahã made many errors on these non-numerical tasks, which makes it difficult to interpret their errors on the numerical tasks. The Pirahã's responses showed *approximately* the right number of objects even in the one-to-one correspondence tasks, suggesting they may not have understood the goal of the tasks. More broadly, Gordon's quasi-experimental design did not provide any direct evidence of a causal relationship between language and thought, and left his data open to a non-Whorfian

interpretation: Perhaps the causal arrow between the Pirahã's numerical language and thought points in the opposite direction. Cultures that have a counting system in language differ from cultures that lack such as system in many ways, making it difficult to isolate the role of language. The Pirahã results suggest that keeping track of large exact quantities is not critical for getting along in Pirahã society. In the absence of any environmental or cultural demand for exact enumeration, perhaps the Pirahã never developed exact number concepts – and consequently, they never developed the words (Casasanto, 2005).

Both of these skeptical concerns were addressed in subsequent studies. A team led by psychologists Michael Frank and Edward Gibson conducted a conceptual replication of Gordon's study (Frank et al., 2008). The Pirahã showed nearly perfect performance on the tasks that could be solved by one-to-one correspondence but still performed poorly when matching quantities greater than 3 in tasks that required exact enumeration – validating Gordon's (2004) claim that the Pirahã lack the ability to mentally represent exact quantities above 3.

This skeptical possibility concerning the role of culture was addressed in a 2011 study led by psychologists Elizabet Spaepen and Susan Goldin-Meadow. They tested numerical competence in Nicaraguan homesigners: deaf individuals who do not have access to any language, oral or manual. The homesigners developed an idiosyncratic set of gestures to communicate, called homesigns, which did not include a count list. Although deprived of language, these homesigning adults are functional members of a numerate society, and had opportunities and motivations to enumerate things exactly. Nevertheless, the homesigners did not consistently show the correct number of fingers when signing about sets greater than three, nor did they reliably match sets of objects greater than three. Like the Pirahã, the homesigners were capable of approximating the target numbers of fingers or objects but not matching them exactly. Thus, even when integrated

into a numerate society, individuals do not spontaneously develop representations of large exact numerosities without access to a conventionalized counting system in language.

A further study led by psychologist Benjamin Pitt (2022) illustrates the tight coupling between individuals' ability to use number words and their exact number concepts. Pitt and colleagues tested members of an Amazonian people, the Tsimané. Like the Pirahã language, the Tsimané language lacks a count list. Still, many Tsimané speakers have some exposure to the count list in Spanish, due to schools established by the Bolivian government. The experimenters found Tsimané adults who had learned the Spanish count list incompletely: Some individuals could accurately recite the numbers up to 12, or 15, or 17 (etc.), but no higher. Pitt and colleagues tested a group of these 'low counters' on number matching tasks like Gordon's (2004). Participants precisely matched numbers of objects, but only up to the highest number that they could reliably count in Spanish (after which they switched to using their approximate number system to match numbers). The philosopher Ludwig Wittgenstein (1922/2004) argued that the limits of one's language determine the limits of one's mental world. Pitt and colleagues' data support this assertion, at least in the domain of exact number.

To date, all of the evidence that number words give rise to exact number concepts is correlational. There has been no training experiment showing that innumerate people can acquire number concepts if they are taught to use number words (setting aside the 'natural training experiment' that has been conducted billions of times as caregivers in industrialized nations teach children the local count list). But the correlational evidence supporting the linguistic bootstrapping of number is strong: Adults who have never been exposed to a count list show reliable use of their parallel individuation system and approximate number system, but no ability to reliably represent exact numbers greater than 3 (Frank et al., 2008). This inability cannot be attributed to lack of

interest in exact enumeration, or a lack of cultural support, since high-functioning adults in a numerate culture still do not develop exact number concepts without a count list (Spaepen et al., 2011). Finally, adults who acquire the count list incompletely can acquire exact number concepts, but only up to the highest number that they can reliably count to in words (Pitt et al., 2022).

3.4.4 Theoretical and practical implications of language-based number concepts

The body of evidence indicating that number words drive the acquisition of exact number concepts addresses two longstanding criticisms of the literature supporting the Sapir-Whorf hypothesis. First, these findings provide evidence for linguistic determinism: the availability of a count list in a language apparently determines its speakers' capacity to represent exact numbers. Second, while acknowledging the existence of some 'Whorfian' effects, skeptics have maintained that there is "no scientific evidence that languages *dramatically* shape their speakers' ways of thinking" (Pinker, 1994, emphasis added). Whether experimental results are 'dramatic' is not scientifically evaluable. But intuitively, the effect of number words could be considered dramatic in two ways. On the level of the individual mind, people exposed to number words can entertain thoughts that are unthinkable by people who lack this linguistic experience: not only thoughts like, "What's the 12th digit of Pi?" but also everyday thoughts like, "He's 6 feet tall," "She's 18 years old," "I earn $15 per hour," and "It's 10:25 pm." On a cultural level, the capacity to represent and manipulate exact numbers is fundamental to the world as we know it, underlying the science and engineering that produced the buildings we live in, the medicines we take, the cars we drive, and the computers at our fingertips. Without exposure to a count list in language, it appears that large exact number concepts would not exist; without large exact number, the modern, technological

world would not exist. As such, large exact number provides an example of a conceptual domain in which language has a dramatic and transformative effect on the mind and the world.

4. Conclusions and future directions

The Sapir-Whorf hypothesis has remained controversial for nearly a century. Whereas the idea that language shapes people's thoughts was widely considered to be untenable throughout much of the 20th century, interest in this hypothesis was renewed around the turn of the 21st century, and a body of research validating this hypothesis began growing. Features of the lexicon or grammar that differ across languages have been shown to influence their speakers' thinking in domains such as *color*, *time*, *space*, *objects*, *causation*, and *music*. Beyond supporting linguistic relativity (i.e., the proposal that habits of language affect habits of thinking), there is now at least one line of evidence for linguistic determinism (i.e., the proposal that one's language enables or precludes some aspects of thinking), in the domain of *number*. People who speak different languages think differently, in a variety of ways. Learning a new language can change the way people think, as can having new features of language introduced in the laboratory, suggesting that language can play a causal role in thought.

On the timescale of cultural evolution, it is likely that concepts generally precede the words that name them, and the existence of the concepts drives the creation of new words. On the timescale of an individual's cognitive development, however, studies suggest that (i.) using words and other features of language can change preexisting concepts, and that (ii.) using words can drive the creation of new concepts and new conceptual domains.

How does language affect thought? The studies discussed here suggest several psychological mechanisms. For example, verbal labels make concepts more accessible in memory

(Brown & Lenneberg, 1954). Using words to label bounded regions of continuous perceptual spaces causes category boundaries in these spaces to be formed or changed (Winawer et al., 2007). Linking two concepts via linguistic metaphors causes these concepts to be more strongly associated in long-term memory (Boroditsky, 2001; Casasanto et al., 2025; Dolscheid et al., 2013). New words invite learners to construct new concepts (Carey, 2004; Gordon, 2004; Frank et al., 2008; Spapen et al., 2011; Pitt et al., 2022). Language activates ordinary cognitive processes; from this perspective effects of language on thought may appear not only explicable but also inevitable.

There are many features of language (e.g., words, morphology, phonology, syntax), and many ways that these features vary typologically across languages; there are many domains of thought, and many cognitive mechanisms by which features of language could shape people's thinking across these domains (e.g., directing attention, strengthening memories, moving category boundaries, creating new concepts). As such, there are likely to be countless effects of language on thought that have yet to be discovered. Exploring ways that language shapes thought can help researchers to learn how minds develop, and to understand both the universal features of human minds and the extent of cognitive diversity.

5. *Works cited*

Alverson, H. (1994). *Semantics and Experience: Universal Metaphors of Time in English, Mandarin, Hindi, and Sesotho*. Baltimore: Johns Hopkins University Press.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology, 43*(1), 1-22.

Boroditsky, L., Schmidt, L.A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner and S. Goldin-Meadow (eds) *Language in Mind: Advances in the Study of Language and Thought*. Cambridge, MA: MIT Press, 61–79.

Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *The Journal of Abnormal and Social Psychology, 49*(3), 454.

Carey, S. (2004). Bootstrapping and the origin of concepts. *Daedalus, 133*: 59–68.

Casasanto, D. (2005). Crying "Whorf!" *Science 307*: 1721–1722.

Casasanto, D. (2008). Who's afraid of the Big Bad Whorf? Cross-linguistic differences in temporal language and thought. *Language Learning*, 58 Suppl 1: 63–79.

Casasanto, D., Fotakopoulou, O., Lucero, C., Pita, R, & Boroditsky, L. (2025). The long and small of it: Language shapes duration estimation in speakers of English and Greek. *Language & Cognition.* DOI: 10.1017/langcog.2025.10055

Chen, J. Y. (2007). Do Chinese and English speakers think about time differently? Failure of replicating Boroditsky (2001). *Cognition, 104*(2), 427-436.

Chomsky, N. (1957). Syntactic structures. Mouton de Gruyter.

Chomsky, N. (1965). Aspects of the Theory of Syntax. MIT Press.

Clark, E. (2003). Languages and representations. In D. Gentner and S. Goldin-Meadow (eds) *Language in Mind: Advances in the Study of Language and Thought*. Cambridge: MIT Press, 17–23.

Clark, H. H. (1973). Space, time, semantics and the child. In T. E. Moore (ed.) *Cognitive Development and the Acquisition of Language*. New York: Academic Press, 27–63.

Cubelli, R., Paolieri, D., Lotto, L., & Job, R. (2011). The effect of grammatical gender on object categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(2), 449.

Dehaene, S. (1999). *The Number Sense: How the Mind Creates Mathematics*. New York: Oxford University Press.

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: psychophysical evidence for linguistic relativity. *Psychological Science, 245*: 613–621.

Fausey, C. M. & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin and Review, 18*: 150–157.

Frank, M. C., Everett, D.L., Fedorenko, E. & Gibson, E. (2008). Number as a cognitive technology: evidence from Pirahã language and cognition. *Cognition, 108*: 819–824.

Gordon, P. (2004). Numerical cognition without words: evidence from Amazonia. *Science, 306*: 496–499.

Hermer, L., & Spelke, E. (1996). Modularity and development: The case of spatial reorientation. *Cognition, 61*(3), 195-232.

Hill, K. C. (1997). Spanish loanwords in Hopi. *Trends in Linguistics Studies and Monographs, 108*, 19-24.

Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition, 118*(1), 123-129.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. Chicago: University of Chicago Press.

Lenneberg, E. H., & Roberts, J. M. (1953). The denotata of color terms. *Paper read at Linguistic Society of America*, Bloomington, Indiana.

Levinson, S. C. (2003). *Space in Language and Cognition: Explorations in Cognitive Diversity*. Vol. 5. Cambridge: Cambridge University Press.

Lohmann, H. & Tomasello, M. (2003). The role of language in the development of false belief understanding: a training study. *Child Development, 74*: 1130–1144.

Lucy, J. A. (1992). *Language Diversity and Thought: A Reformulation of the Linguistic Relativity Hypothesis*. Cambridge: Cambridge University Press.

Malotki, E. (1983). *Hopi time: A linguistic analysis of the temporal concepts in the Hopi language* (Vol. 20). Walter de Gruyter.

Majid, A., Bowerman, M., Kita, S., Haun, D. B., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences, 8*(3), 108-114.

Mead, M. (1928). *Coming of Age in Samoa: A psychological study of primitive youth for Western civilization*. William Morrow & Company: New York.

Miles, L. K., Tan, L., Noble, G. D., Lumsden, J., & Macrae, C. N. (2011). Can a mind have two time lines? Exploring space–time mapping in Mandarin and English speakers. *Psychonomic Bulletin & Review, 18*, 598-604.

Núñez, R. E., & Sweetser, E. (2006). With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science, 30*(3), 401-450.

Papafragou, A., J. Hulbert and J. Trueswell 2008. Does language guide event perception? Evidence from eye movements. *Cognition, 10*: 155–184.

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition, 87*(1), B35-B45.

Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York: Harper.

Pitt, B., Gibson, E., & Piantadosi, S. T. (2022). Exact number concepts are limited to the verbal count range. *Psychological Science, 33*(3), 371-381.

Pullum, G. K. (1991). *The Great Eskimo Vocabulary Hoax and Other Irreverent Essays on the Study of Language*. Chicago: University of Chicago Press.

Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22*(3), 341-365.

Sapir, E. (1929). The status of linguistics as a science. *Language, 5*: 207–214.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking." In J. J. Gumperz and S. C. Levinson (eds) *Rethinking Linguistic Relativity*. Cambridge: Cambridge University Press, 70–96.

Song, S., Miller, K. D., & Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience, 3*(9), 919-926.

Spaepen, E., Coppola, M. Spelke, S., Carey, E., & Goldin-Meadow, S. (2011). Number without a language model. *Proceedings of the National Academy of Sciences, 10*: 3163–3168

Spelke, E. S. (2000). Core knowledge. *American Psychologist, 55*(11), 1233.

Watson, J. B. (1928). *The ways of behaviorism*. Harper & Brothers.

Whorf, B. L. (1939/1956). The relation of habitual thought and behavior to language. In J. Carrol (Ed.) *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. MIT Press.

Whorf, B. L. (1940/1956). Science and linguistics. In J. Carrol (Ed.) *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. MIT Press.

Whorf, B. L. (1941/1956). Languages and logic. In J. Carrol (Ed.) *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. MIT Press.

Winawer, J., Witthoft, N., Frank, M.C., Wu, L., Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Science, 104:* 7780–7785.

Wittgenstein, L. (1922/2004). *Tractatus Logico-Philosophicus.* Project Gutenberg eBook.

6. *Suggestions for further reading*

Everett, D. (2010). *Don't sleep, there are snakes: Life and language in the Amazonian jungle*. Pantheon Books.

Levinson, S. C. (2003). *Space in Language and Cognition: Explorations in Cognitive Diversity*. Vol. 5. Cambridge: Cambridge University Press.

Whorf, B. L. (1956). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. Edited by John B. Carroll. Cambridge: MIT Press.