1 What Are the Key Concepts?

The question of whether the language we speak shapes the way we think has generated extensive debate in recent decades. The study of how language influences thought, also known as linguistic relativity (Whorf, 1956), has recently received renewed interest as a number of new research paradigms have evolved that allow addressing the interplay between language and thought empirically.

Experimental evidence suggests that cross-linguistic differences in linguistic encoding can 'augment certain types of thinking' (Wolff & Holmes, 2011, p. 253), such as attention, recognition memory, visual discrimination, sorting and categorization, in a flexible and context-dependent manner. For instance, cross-linguistic differences in colour vocabularies can cause differences in colour categorization, indicating that language effects are profound in the sense of affecting even basic 'categorical perception' (i.e., faster or more accurate discrimination of stimuli that straddle a category boundary, Regier & Kay, 2009, p. 439). However, such linguistic relativity effects are vulnerable to short-term manipulations, such as recent linguistic priming (Montero-Melis et al., 2016), the language of instruction (Athanasopoulos, Bylund et al., 2015) and verbal interference (Gennari et al., 2002).

Given the great complexities of language-and-thought research, in Section 1, we introduce a number of key terms and concepts surrounding the notion of the language–thought interface. Specifically, we are going to discuss what thought is, where we can find it and in what form it is manifested. We will then go over current views of when and where language effects on thought are most likely to arise. This helps to sketch out the cognitive mechanisms at play for linguistic relativity effects and paves the way for further discussions of the language-and-thought interface in speakers of more than one language.

1.1 What Is the Language-and-Thought Interface?

The theoretical basis of the language-and-thought interface is grounded in one of the most extensively debated theories, namely the linguistic relativity hypothesis (LRH), also known as the Sapir–Whorf hypothesis (Whorf, 1956). The LRH postulates that language and thought are interrelated, and people who speak different languages embody different world views depending on the linguistic categories made available in those languages. For example, when expressing the duration of time, speakers of Swedish and English, who prefer distance-based metaphors and describe time as 'long' and 'short' (Alverson, 1994; Evans, 2004), represent the passing of time differently compared to speakers of Greek and Spanish, who prefer amount-based metaphors and

describe time as 'big' and 'small' (Casasanto et al., 2004; Lakoff & Johnson, 1980).

While the idea that language is related to thought extends a long way back to the early days of Western philosophy (see Lucy, 1997, 2016, for a historical review) – for instance, von Humboldt viewed language and thought as an inseparable unit with each language giving its speakers a particular 'worldview' (von Humboldt, 1963, p. 60) – this issue gained its most prominence via the work of Edward Sapir and Benjamin Lee Whorf. In Whorf's words,

> We cut nature up, organize it into **concepts**, and ascribe significances as we do, largely because we are parties to an agreement to organize it in this way-an agreement that holds throughout our speech community and is codified in the patterns of our language. [. . .] We are thus introduced to a new principle of relativity, which holds that all observers are not led by the same physical evidence to the same picture of the universe, unless their linguistic backgrounds are similar, or can in some way be calibrated. (Whorf, 1940, pp. 213–214)

The basic tenet of the LRH is that languages 'carve up' the world in different ways. Thus, if the language people speak constrains them to attend to the external world in certain ways, speakers of different languages will develop distinct views and concepts of the reality in ways that reflect language-specific properties.

> The linguistic relativity principle . . . means . . . that users of **markedly different grammars** are pointed by **their grammars** toward different types of observations and different evaluations of externally similar acts of observation, and hence are not equivalent as observers but must arrive at somewhat **different views of the world**. (Whorf, 1940/1956, p. 221)

To be more specific, Whorf (1956, p.158) clearly explains each notion and puts emphasis on the effects of 'grammars' and 'different views of the world'. By 'grammars', Whorf means habitual lexical and grammatical patterns of a language, such as 'lexical, morphological, syntactic, and otherwise systemically diverse means coordinated in a certain frame of consistency'. 'Different views of the world', on the other hand, is typically understood as 'thought' or 'concepts' that modulate speakers' habitual or routinized ways of conceptualizing, perceiving and classifying reality.

Following Lucy (1992a, 1992b), the definitions of 'concepts' and 'thought' in present-day psycholinguistic research are typically operationalized as a wide array of non-linguistic (or non-verbal) behaviours and mental processes, including attention, reasoning, perception recognition memory, problem-solving, sorting and categorization. These processes are non-verbal in nature because they do not involve or elicit overt language comprehension or production, but

manifest certain forms of cognitive or perceptual responses to given stimuli (Athanasopoulos & Bylund, 2020; Bylund & Athanasopoulos, 2014b; Casasanto, 2008; Gallistel, 1989). This line of reasoning also resonates with Pavlenko's (2005, p. 435) definition of concepts, which refers to 'mental representations that affect individuals' immediate perception, attention, and recall and allow members of specific language and culture groups to conduct identification, comprehension, inferencing and categorization along similar lines'.

1.1.1 Does Language Determine Thought?

Since its formulation, the LRH has taken various forms (for an evaluation of different models, see Wolff & Holmes, 2011). A strong view of it is known as linguistic determinism, which holds that language shapes or determines thought (Brown & Lenneberg, 1954). According to Brown and Lenneberg (1954), 'languages are moulds into which infant minds are poured' (p. 454). This language-as-mould metaphor, as pointed out by Casasanto (2008, 2016), has two major flaws. First, language does not necessarily shape thought permanently or at only one point in time (i.e., early childhood) during one's entire cognitive development. Second, language is not the sole shaper of thought. In line with Casasanto's influential argument, Wolff and Holmes (2011) also remind us that although language triggers thought, thinking is possible without language. In this view, linguistic determinism overexaggerates the shaping role that language has. In fact, research from the cognitive sciences has indicated that the relationship between thought and the world is much tighter than that between thought and language, with plenty of evidence showing that differences between languages are much more diverse than differences observed in people's mental representations (Casasanto, 2016; Lucy, 1992a, 1992b; Munnich, Landau, & Dosher, 2001; Regier & Kay, 2009; Roberson, 2005; Wolff & Malt, 2010).

Despite having been hotly debated for more than half a century, language determinism has difficulty in holding ground as no empirical evidence has ever been found to support this radical claim. In contrast, an increasing amount of research has illustrated that linguistic relativity effects, rather than rigidly shape one's world view, tend to mediate or affect aspects of cognition in a flexible and context-dependent manner (Casasanto, 2008; Levinson, 2001; Trueswell & Papafragou, 2010; Wolff & Holmes, 2011). For example, in the domain of colour, recent studies show that cross-linguistic differences in colour naming affect the discrimination of colour (Athanasopoulos et al., 2010; Regier & Kay, 2009; Roberson, 2005; Thierry et al., 2009; Winawer et al., 2007). For instance,

Winawer et al. (2007) reported that obligatory colour distinctions in Russian between *golouboy* (light blue) and *siniy* (dark blue) affected one's processing speed in colour discrimination. Speakers of Russian, who have two basic lexical items for blue (*golouboy*/*siniy*), responded much faster when matching two colours belonging to different categories (i.e., one is golouboy and the other is siniy) than within the same category (i.e., both are bolouboy or siniy). However, such patterns were not observed in speakers of English, who only have one basic lexical item for blue. In a similar vein, Thierry et al. (2009) examined how colour category boundaries affect categorical perceptions of colour in Greek and English speakers using event-related potential (ERP) techniques. It was found that speakers of both Greek and English were perceptually aware of the distinctions between two shades of blue and green, as indicated by their brain activation patterns. But Greek speakers, who have two basic colour items for light and dark blue (i.e., *ghalazio* and *ble*), but only one basic item for light and dark green (i.e., *prasino*), displayed greater brain activation for blue than for green contrasts, thus showing that language categories affect colour discrimination performance.

On the other hand, turning to the domain of motion, which refers to 'a situation containing movement of an entity or maintenance of an entity at a stationary location' (Talmy, 1985, p. 60), Papafragou et al. (2002) reported that although speakers of English (satellite-framed language) and Greek (verb-framed language) differed significantly in how motion is talked about (i.e., English: a man walking across the street; Greek: a man crossing the street (walking)), their categorical preferences for manner and path were far more similar than their naming patterns.

In summary, the overall findings suggest that the effect of language on cognition is not a simple 'yes-or-no' question. The fundamental issue here is to discover what aspects of language tend to affect what dimensions of thinking and in what ways (Bylund & Athanasopoulos, 2014b; Casasanto, 2016).

1.1.2 What Are the Main Controversies?

Over the years, the LRH has sparked much controversy in the disciplines of Linguistics, Psychology, Philosophy and Anthropology (for critical discussions, see Casasanto, 2008, 2016; Lucy, 1997, 2016; Pinker, 1994). On the one hand, criticism ~~on~~ of the LRH partly comes from the strong version of linguistic determinism, since it oversimplifies the intricate connection between language and cognition. As pointed out by Pavlenko (2011, p.19), interpreting the LRH as a simple 'yes-or-no dichotomy' is a misinterpretation of Whorf's original concepts. In fact, the determinism view is a later invention introduced

by those who attempted to reformulate the ideas but lost their original arguments in translation (Pavlenko, 2011, 2016).

On the other hand, at the opposite extreme is the 'universalist' position, in which thought is said to be free, universal and entirely independent of language (Pinker, 1994; see Casasanto, 2008; Thierry, 2016, for further discussions and critical evaluation). Being the best-known criticism of linguistic relativity, the rationale behind this approach is the Universal Grammar (Chomsky, 1975), which claims that human cognitive behaviours are guided by universal perceptual biases and not subject to language-specific properties. Although language-specific properties can reflect some facets of our cognitive functions, they neither mould nor guide thought (1975, p. 4). Pinker further suggests that 'language is not necessary for concept acquisition nor does it "pervade[e]" thought' (Pinker, 1994, p. 17). Under this view, the notion that 'differences among languages cause differences in the thoughts of their speakers' is 'wrong, all wrong' (Pinker, 1994, p. 57).

The universalist approach has triggered numerous debates concerning the link between language and the rest of the mind. In a later explanation, Pinker (1994) proposes that 'the idea that thought is the same as language is an example of what can be called a conventional absurdity [.]' (p. 57). However, as highlighted by Casasanto (2008), most of the critiques from the universalist camp centre around the so-called Orwellian claim that 'the idea that thought is the same as language' (Orwell, 1949), rather than the Whorfian effect (i.e., whether language shapes thought). As a result, rejecting the language as language-of-thought assumption does not mean that we are ready to accept the opposite view and consider language–thought interdependence as an alternative option.

In fact, several groundbreaking studies from cognitive neuroscience and brain neurophysiology have challenged the universal dominance of human cognition and view language as an essential and indispensable part of the human mind (Athanasopoulous et al., 2010; Boutonnet et al., 2013; Thierry et al., 2009). This perspective is well reflected in Thierry (2016), who further suggests that it is misleading and essentially meaningless to separate language from the rest of cognitive general abilities, and 'thinking that language may be entirely disconnected from thought is an example of what deserves to be called *reductio* ad *absurdum*' (p. 691). To sum up, the main controversies regarding linguistic relativity effects are situated at two extremes of a conceptual continuum. While we do not find support for these two opposing views, there is converging evidence that language can affect cognition via numerous mechanisms (see Wolff & Holmes, 2011, for an overview). We will focus on different mechanisms via which language affects cognition in the following sections.

1.1.3 Contemporary Approaches to Language-and-Thought Research

As noted earlier, neither the universalist view nor a strong view of linguistic determinism can successfully unravel the mechanism underlying the relationship between language and thought. With the development of multidisciplinary research in the twentieth century, the LRH received renewed interest after the dominance of the universal-based approach. Contemporary approaches to language–thought research (also known as the 'neo-Whorfian' approach) adopt a multidisciplinary perspective and emphasize the need to implement both linguistic and non-linguistic paradigms when addressing the language-and-thought debate (Bylund & Athanasopoulos, 2014b; Levinson, 2003; Lucy, 1997, 2014, 2016; Majid et al., 2004). For instance, contemporary researchers have begun to place psycholinguistic methods at the centre of testing and try to operationalize the Whorfian question using modern cognitive theories that precisely characterize the nature and complexity of cognitive effects (Lucy, 1992a, 1992b; Lupyan, 2012; Slobin, 1996). Furthermore, recent advances in the fields of neurophysiology and the cognitive neurosciences have provided researchers with new opportunities to discover the neural correlates of language effects on cognition, thus providing more nuanced evidence for the complexities involved in language and cognitive processing (see Athanasopoulos & Casaponsa, 2020, for a recent review).

To be more specific, by directly utilizing a variety of behavioural measures, such as triad-matching, recognition memory, attention allocation, as well as reaction times (Levinson, 2001, 2003; Lucy & Gaskins, 2001, 2003; Papafragou et al., 2008; Regier & Kay, 2009; Roberson & Davidoff, 2000), together with neurophysiological techniques, such as eye-tracking, ERPs and functional MRI (Athanasopoulos et al., 2010; Boutonnet et al., 2013; Flecken et al., 2015a; Thierry et al., 2009), the interface between language and attention has been examined in a wide array of cognitive domains. For instance, findings from the domain of colour suggest that speakers with different word labels for colours are found to be more efficient in colour recognition (Franklin et al., 2008; Gilbert et al., 2006; Regier & Kay, 2009; Roberson et al., 2008; Thierry et al., 2009; Winawer et al., 2007). Cross-linguistic differences across languages have also been observed in how people think about objects and substances (Ameel et al., 2005; Imai & Gentner, 1997; Pavlenko & Malt, 2011), as well as more abstract conceptual categories such as time (Boroditsky et al., 2011; Boroditsky et al., 2003; Casasanto & Boroditsky, 2008), number (Athanasopoulos, 2006; Cook et al., 2006; Lucy, 1992a; Lucy & Gaskins, 2001, 2003), gender (Bassetti & Nicoladis, 2016; Bender et al., 2018; Sato & Athanasopoulos, 2018), spatial frames of reference (Levinson, 2001, 2003; von Stutterheim et al., 2017) and

motion events (Flecken et al., 2015a; Gennari et al., 2002; Montero-Melis & Bylund, 2017; Trueswell & Papafragou, 2010).

The prevailing question 'Does the language we speak influence the way we think?' can then be replaced by a battery of more specific investigations, such as when and where are the language effects on cognition mostly likely to appear? What is the exact cognitive mechanism that gives rise to such effects? And how can we most precisely uncover the nature of such effects? The answers to these questions will provide us with more precise definitions of aspects of language and cognition and advance our understanding of the role of language within cognition.

1.2 Thinking for Speaking: Where Is It?

The notion of thinking-for-speaking (TFS) was proposed by the psycholinguist Dan I. Slobin (1987, 1996, 2000, 2003) as an influential version of the LRH (Casasanto, 2015), arguing that the activity of thinking takes on a particular quality when it is employed in the activity of speaking (Slobin, 1996, p. 76). Specifically, when Slobin talks about 'thinking for speaking', he focuses on the effect of language on thinking that is conducted during the processes of speaking, writing, translating or remembering. From this perspective, thinking is 'a special form of thought that is mobilised for communication' (Slobin, 1996, p. 6) and its effect is limited to online processes only.

More specifically, TFS postulates that language channels one's attention. When people are involved in language-induced activities, such as comprehension or speech production, they need to pick those elements that (1) fit some conceptualization of the event and (2) are readily encodable in language (Slobin,1987, p. 435). As a consequence, the linguistic constraints of different languages may guide speakers to attend to specific details of information when talking about them. The crucial difference between the LRH and TFS, therefore, is that the former emphasizes the language effects on habitual thought regardless of whether language is being in use or not, while the latter focuses on thinking patterns during active language use.

According to Slobin (1991)one way to investigate 'thinking for speaking' is by focusing on a child's first language acquisition. When children acquire a native language, they might learn particular ways of thinking (p. 2). Research along these lines is interested in exploring whether child speakers of different languages exhibit language-specific thinking patterns and when during the course of L1 development TFS effects start to appear (Allen et al., 2007; Berman & Slobin, 1994; Choi & Bowerman, 1991). Another way to gain insights into 'thinking for speaking' is by looking at L2 learners, concentrating

on the questions of (1) whether language patterns acquired in childhood are 'resistant to restructuring in adult second language acquisition' (Slobin, 1996, p. 89), and (2) what difficulties learners may encounter during the acquisition of thinking patterns associated with the L2.

The TFS hypothesis has mainly been investigated in the motion domain, starting with observations of speakers' speech and gesture patterns. In terms of expressions of motion, one influential framework within this line of research is based upon Talmy's (1985, 2000) typological distinctions between so-called satellite- and verb-framed languages. For example, when expressing the mundane event 'A boy walks up a hill', speakers of satellite-framed languages (S-languages) such as English and German, encode the manner of motion ('walk') in the main verb, whereas manner-of-motion information can be omitted in other languages. Talmy's typological framework has turned out to be particularly useful in cross-linguistic comparisons across the world's most spoken languages. By using a wide range of speech elicitation methods (i.e., spontaneous speech and narrative production), studies examining the speech and gestures of children and adult speakers of various languages (Duncan, 2001; Hickmann & Hendriks, 2006, 2010; Hickmann et al., 2009; Kita & Özyürek, 2003; McNeill, 1997, 2000; Montero-Melis & Bylund, 2017; Özyürek et al., 2005) have reported apparent cross-linguistic differences in how speakers think, speak and gesture about motion.

While substantial evidence suggests that speakers of different languages select and organize information in language-specific ways (Berman & Slobin, 1994; Slobin, 2003, 2006; von Stutterheim et al., 2017; von Stutterheim & Nüse, 2003), recent studies have questioned whether differences between languages can be equated with differences in thought processes (Athanasopoulos & Albright, 2016; Casasanto, 2008, 2016; Lucy, 2014, 2016). As pointed out by many scholars, using linguistic data alone may run the risk of circular reasoning (i.e., language-on-language effect), since the only evidence that people who talk differently also think differently is that they talk differently (Casasanto 2008, p. 67). To gain a better understanding of the cognitive implications of language-specific features in human thinking processes, an increasing number of studies within the TFS paradigm have started to combine speech production with a variety of dynamic measures to capture the mental processes concurrent with verbal production.

These methods include experimental paradigms using multimodal tasks, such as attention, recognition memory and categorization, often coupled with co-verbal behaviours that involve gestures (Brown & Gullberg, 2008; Cadierno, 2008; Stam, 2015), eye movements (Flecken et al., 2014; von Stutterheim et al., 2012), reaction times (Ji & Hohenstein, 2018; Wang & Li, 2021b) and ERPs

(Athanasopoulos et al., 2010; Flecken et al., 2015a). For example, using an eye-tracking paradigm, Papafragou et al. (2008) explored how Greek and English speakers directed their visual attention to different components of motion events (i.e., manner and path of motion) in the process of speech preparation. Results showed that compared with English speakers, Greek speakers were more likely to prioritize their attention to path over manner when preparing for speech. However, such language-specific effects subsequently disappeared when participants simply watched the motion scenes freely. In a similar vein, von Stutterheim et al. (2012) explored how the presence and absence of grammatical aspect in the target language influence the extent to which speakers attend to different components (i.e., the goal or the ongoing phase) of motion events. Grammatical aspect is a linguistic category that denotes the internal temporal property of a situation (Comrie, 1976; Dahl, 2000). For example, in English, grammatical aspect is systematically encoded on the main verb, and there is an obligatory distinction between the ongoing (i.e., A man is crossing the street, imperfective/progressive aspect) and the completed (i.e., A man has crossed the street, perfective aspect). However, in other languages, such as German and Swedish, there is no such grammatical device to convey this contrast.

Combining language production with attention allocation and recognition memory, von Stutterheim et al. (2012) reported that speakers of languages that provide obligatory grammatical means to convey aspectual contrasts, or aspect languages (i.e., English, Arabic and Spanish) tended to mention end points less frequently compared with speakers of languages that lack obligatory grammatical means to denote such contrasts, or non-aspect languages (i.e., German, Dutch and Czech). At the same time, speakers of aspect languages directed less attention to end points and stored less information about them in their working memory. In addition, visual attention patterns provided a more nuanced picture, that is, speakers of aspect languages also looked at event end points at a later point than speakers whose languages do not contain such grammatical device during speech planning. The findings thus indicate that looking at co-verbal behaviour such as visual attention provides us with a unique window into real-time processing in the preparation of describing an unfolding event (Athanasopoulos & Casaponsa, 2020).

Using ERP techniques, Flecken et al. (2015a) studied the influence of grammatical aspect on the perceptual processes of event construal in speakers of English and German. As noted earlier, speakers whose languages have grammatical aspect (i.e., English) are more likely to attend to the ongoing phase of an event (i.e., A man is walking along a street) than speakers whose languages lack the progressive aspect (i.e., German) (von Stutterheim et al., 2012). In the

experiment, participants were asked to perform a matching task with their ERPs being recorded. In each trial, participants watched a one-second animated prime showing basic motion events, in which a dot travelled along a trajectory (curved or straight) towards an end point of a geometrical shape (square or hexagon). But the end point was never reached. Then participants were engaged in a picture matching task in which the target animation was followed by a picture symbolizing event end points or trajectories in four different conditions: a full-match condition (5 per cent) where both end point and trajectory matched the target (i.e., a dot moving along a curved trajectory towards a square), a full-mismatch condition (75 per cent) (i.e., a dot moving along a straight trajectory towards a hexagon), an end point match condition (10 per cent) (i.e., a straight line) and a trajectory match condition (10 per cent) (i.e., a hexagon). Participants were instructed to respond only to the full-match condition. Results showed that speakers of German displayed larger P3 amplitude (an ERP component used for reflecting conscious processes involved in attentional processing) in end point match conditions than in trajectory match conditions, while English speakers did not display any differences. The authors therefore concluded that cross-linguistic differences in grammatical properties affected lower-level processing, and speakers of different languages automatically allocated their attention to the linguistic elements highlighted by grammar.

Using a triad-matching paradigm, Gennari and colleagues (2002) took the first step within the TFS paradigm and investigated whether cross-linguistic diversities in linguistic encoding moved beyond verbal behaviour and affected English (S-language) and Spanish (V-language) speakers' recognition of and categorical preferences for manner and path. Speakers of English and Spanish were allocated to one of three conditions: a 'naming first' condition, during which participants described all motion videos prior to recognition and similarity judgements; a 'free' condition where participants watched motion in silence; and a 'shadow condition' in which participants were instructed to repeat aloud nonword syllables while watching the videos. Results showed that while shadowing led to an overall decrease in path-congruent selections in both English and Spanish, only Spanish speakers selected significantly more path-congruent choices after the 'naming first condition' compared with the 'shadow condition'. In a more recent study, Wang and Li (2021b) coupled the triad-matching paradigm with reaction times and extended the domain of interest to early Cantonese–English bilinguals whose language pairs do not exhibit contrastive typological differences. In the 'naming first' condition, the bilinguals were randomly allocated to either a Cantonese- or English-speaking condition during which they had to verbalize all motion videos in the target language prior to similarity judgements. In the 'shadowing condition', participants had to