The decreasing Whorf-effect: A study in the classifier systems of Mandarin and Thai

Fabian Bross[1], **Philip Pfaller**

Received Feb 14th, 2012, accepted June 20th 2012, published June 22nd 2012

The goal of this study was to test a weak form of the Sapir-Whorf Hypothesis dealing with one of the biggest unsolved questions in linguistics: Does language affect the way we think? Grammatical systems in the world's languages differ in many aspects. Unlike English or German many languages group nouns on the basis of noun classifiers. Recently research has adressed the question if these linguistic categories built up by classifier systems influence non-linguistic thought. In this paper we studied Mandarin Chinese and Thai—two languages with classifier systems. Although both are classifier languages they categorize objects in different ways. We tested if these system differences lead to different similarity judgements of objects in a non-linguistic rating task (participants had to rate the similarity of picture pairs). In contrast to previous studies we suprisingly observed no difference in categorization. It seems that the so-called Whorf effect, i. e. that language affects the way we perceive and categorize the world, diminishes rapidly over the time speakers are exposed to a different language system such as, in this case, German.

1 Introduction

"With every new language you learn, you get a new soul."
Czech proverb

One of the main questions of cognitive science is how humans categorize objects. One of the key issues of linguistic research therefore is the relationship between human cognition and linguistic categories. In the world's languages there are many grammatical categories that classify nouns in several groups like gender grammar systems, which typically have two (masculine/feminine) or three (masculine/feminine/neuter) categories. Another grammatical category are classifier grammar systems grouping nouns into semantic categories. The question if and how classifier grammar systems could impact human categorization has been addressed by several researchers (e. g. [1-7]). This research indicates that linguistic categories have an influence on how speakers conceptualize the world—an idea which is strongly associated with Edward Sapir and Benjamin Lee Whorf. Nevertheless, until today the question if and how language affects the way we think is highly contested in scientific literature.

2 The Sapir-Whorf Hypothesis

The origin of linguistic relativity is found in the works and thoughts of Wilhelm von Humboldt, Franz Boas, Edward Sapir and Benjamin Lee Whorf. In this part we will just handle the two eponymous researchers in detail.

In *Language, Thought, and Reality*, the collection of Whorf's work released after his death, he calls Franz Boas the first scientist in history who showed how language could be analyzed *sui generis* in a scientific way without forcing it to the classic categories (cf. [8], p. 114). The origin of Whorf's concept that people who speak languages with very different grammars are lead by these grammars to different observations and different evaluations of these observations, can be found in Boas' writings:

> Languages differ not only in the character of their constituent phonetic elements and sound-clusters, but also in the groups of ideas that find expression in fixed phonetic groups. ([9], p. 24)

Whorf ([8], p. 114) considers Edward Sapir's *Language*, a book published in 1921, a precursor to a new linguistic era. In this book Sapir shows the importance of linguistic studies on thought. There he defines the term language as

> a purely human and noninstinctive method of communicating ideas, emotions, and desires by means of a system of voluntarily produced symbols ([28], p. 8).

According to Sapir, the relationship between these symbols does not consist of phonemes and meanings, but of mental categories and physical actions for phoneme generation. Sapir says that there is no thinking without language and he points out that his concept of language is not that of speech (cf. [10], p. 136).

[1] fisimatenten@gmx.de

Whorf adopts Sapir's general view that the perceivable world is created subconsciously by the language use of a group of speakers. This leads him to propose that each speech community has its own interpretations when making certain experiences (cf. [8], p. 74). Whorf believed that we think of things in a certain way because of the language we speak (cf. [8], p. 9). Whorf wrote that linguistic processes are essential to reaching consensus or agreement ([8], p. 11).
It is important to point out that Whorf, like Sapir, does not equate language with thinking. Sapir wrote:

> At best language can but be the outward facet of thought on the highest, most generalized level of symbolic expression ([28], p. 15).

Sapir compares language to an instrument and thought to a product. With a product one can refine the instrument ([11], p. 24). Thus the function of language is not to label the thought. Rather language and thought are two sides of one and the same process. Language influences our way of thinking. While neither Sapir nor Whorf formulated a scientifically testable hypothesis, Brown ([12], p. 128) stated the following two hypotheses:

1. Structural differences between language systems will, in general, be paralleled by non-linguistic cognitive differences, of an unspecified sort, in the native speakers of the two languages.

2. The structure of anyone's native language system strongly influences or fully determines the worldview she will acquire as she learns the language.

Brown was influenced by Eric Lennebergs article *Cognition in Ethnolinguistics* in 1953 in which he criticizes that "the basic assumption that language affects non-linguistic behavior derives from an inspection of linguistic facts" ([13], p. 464). These so-called *weak* and *strong* forms of the linguistic relativity hypothesis served as starting point for the work of many later researchers.
To conclude this section we quote Lucy giving a simplified definition of the linguistic relativity hypothesis in its weak form:

> In its most elementary form, the linguistic relativity hypothesis posits that diverse languages influence the thought of those who speak them ([1], p. 263).

3 Classifiers and categorization

Classifiers group nouns on the basis of certain characteristics of the referred entities. According to Allan ([14], p. 285), a numeral classifier is a morpheme which takes "some salient perceived or imputed characteristic of the entity to which the associated noun refers." Beside a few other uses[2] numeral classifiers (henceforce referred to as classifiers) are syntactically obligatory when counting (cf. [14] or [15], p. 43). They conjugate with numerals or determiners and constitute noun phrases (cf. [16], p. 589).
Classifiers are used in many languages, especially in Asia, America, Africa and Oceania (e.g. [14], p. 285 or [17], p. 1).[3] Zhang ([15], p. 44) concludes that classifiers employ categorizations independent of the classifier language with "roughly the same set of parameters [...], most notably, animacy, shape, function, consistency, and size" (see also [18], p. 404 or [14], p. 307). Inglis ([19], p. 237) notes that a classifier "categorizes a noun on the basis of some schematic feature or shape specification intrinsic to the noun. A measure word does not categorize the noun but quantifies it on the basis of some standard of measurement, such as a cup" (see also [14], p. 304).
From considerations like Lakoff's ([20], p. 110,) for whom linguistic categories and cognitive categories are part of one system, the question whether or not the categories built by classifiers have cognitive consequences emerges: "One possibility is that classifier systems provide an alternative organization of object concepts [...]" ([6], p. 486). For Lakoff ([21]), the category memberships induced by classifiers are not arbitrary ones but motivated enhancements from central categories. In his research on the Japanese classifier *hon*, which is used for thin, rigid, and long objects like pencils or trees but also for hits in baseball or radio and tv programs, he concludes that such enhancements may be motivated, not arbitrary, because: "They do not have anything in common with long, thin rigid objects, but it makes sense that they might be classified in the same way" ([21], p. 26).[4]

4 Previous studies

If and how classifiers systems have an impact on human categorization has been recently addressed by several studies. For example [4] found such effects for similarity judgement, classification, memory and choice tasks with written stimuli referring to concrete objects. Chinese speakers presented with words referring to objects sharing the same classifier rated these as being more similar than English speakers did. In a similar study, [3] investigated if speakers of classifier languages use linguistically mediated (by categories built by classifiers) conceptual knowledge in a non-linguistic similarity rating task. They presented speakers of English and Chinese with picture pairs where one half consisted of pairs sharing the same classifier in Chinese and the other half of pairs not sharing the same classifier. Whereas the pairs not sharing the same classifier were not rated differently by English and Chinese speakers, Chinese speakers rated the pairs sharing the same classifier as more similar than the speakers of English did.
Saalbach & Imai ([6]) conducted several experiments (a forced choice categorization task, a similarity rating task,

[2] In Chinese numeral classifiers are for example also necessary in demonstrative expressions (e.g. [22], p. 104).
[3] [30] offers a very detailed overview of the classifier systems in the world's languages.
[4] For a description of the selection of Japanese classifiers according to different perspectives see [29].

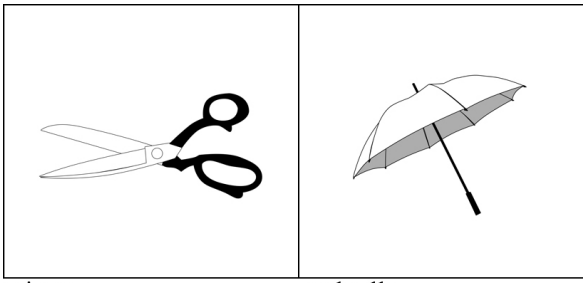
and a priming experiment) with Chinese and German speakers. Like [4], they provide evidence that speaking a classifier language has an impact on cognitive tasks.

These studies show that there is evidence for a weak form of the Sapir-Whorf Hypothesis, i. e. that linguistic categories influence human categorization. In a non-linguistic similarity judgment task, Kuo & Sera (2007) found that Chinese speakers were not only influenced by classifiers compared to English speakers, but that this effect increased with language exposure [5]. With our research we wanted to test the weak form of the Sapir-Whorf Hypothesis.

5 Our research

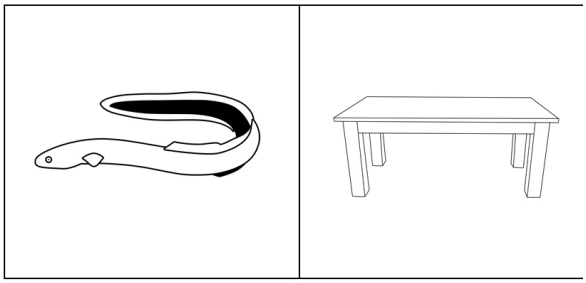
5.1 Pretest

In a short picture-naming task we investigated if our pictures would evoke the intended targets. We showed 39 pictures to ten German students (four female, six male) with an average age of 23 years and instructed them to name the shown objects. There were just two "wrong" answers. One participant named a "pistol" "weapon". Another participant named an "airplane" "Boeing". Because these were—in terms of prototype theory—descriptions above and below the basic level respectively and would not have an influence on the choice of classifier, 100 % of the pictures were named as intended.



scissors
Mandarin classifier: bǎ
Thai classifier: lêm (เล่ม)

umbrella
Mandarin classifier: ba
Thai classifier: k^{h}an (ก้น)



eel
Mandarin classifier: tiáo
Thai classifier: tua (ตัว)

table
Mandarin classifier: zhang
Thai classifier: tua (ตัว)

Figure 1: Examples of black and white illustrations used in the experiment

5.2 Experiment

As in [4], the goal of our experiment was to determine the effects of different classifier systems on cognition. For this purpose we chose Thai and Mandarin speakers as subjects. Thai and Mandarin are both classifier languages but with different systems. Unlike [4], we chose participants who were indeed native speakers but lived Germany so that they were highly exposed to a non-classifier language.

We compared Mandarin and Thai speakers in a non-linguistic similarity judging task. The participants were asked to judge the similarity of picture sets on a rating scale. We predicted that the participants would rate the picture pairs sharing the same classifier in their mother tongue as more similar than the pairs which did not.

5.2.1 Participants

The participants were 18 Mandarin and 21 Thai speakers. The speakers of Mandarin were all university students who had lived in Germany for an average period of one year. They had an average age of 25. The speakers of Thai were all students of a language school (learning German) with the exception of one student at university. They had an average age of 30 and had been living in Germany for an average period of five years. All participants were paid for taking part in the experiment.

5.2.2 Design and Materials

The stimuli were 38 picture pairs: half of the pairs referring to concrete objects sharing the same classifier in Thai and the other half referring to concrete objects sharing the same classifier in Mandarin. An example is shown in Figure 1. The top picture pair refers to nouns sharing the same classifier in Mandarin; the pair below refers to nouns sharing the same classifier in Thai.

The stimuli were chosen by the criteria that the objects shown together in a pair should have as little a relation—taxonomic and thematic—as possible.[5] The arrangement of the pairs was randomized and the experiment conducted with Psyscope (cf. [23]).

5.2.3 Procedure

The participants were tested individually. They were presented with instructions in their native language and in English on the computer screen. Then they had to rate the similarity between the pictures they were shown on a 1-9 rating scale: 1 meaning "no similarities" and 9 being "very similar". There was no time limit and they were instructed to follow their intuition. The complete English instructions are given below.

> Welcome to our experiment!
> You will now be presented with several pairs of pictures, which resemble objects that are not similar to one another. Please intuitively rate

[5] This was done because we did not want the participants to really judge the object's similarity but to test if the grammatical properties induced by the particular classifier system would have an impact on their judgements.

> the similarity of these objects on a scale from 1 to 9 in spite of their differences. 1 means "no similarity at all" and 9 means "very similar". Thank you! Please press any key to start.

After the experiment all participants filled out an information sheet that asked their age, gender, native language and how long they have been in Germany (in months).

5.2.4 Results and Discussion

Contrary to our predictions there was no significant difference between the ratings of Mandarin and Thai speakers. The mean ratings are shown in the following table.

	Mean ratings:	
	Mandarin speakers:	**Thai speakers:**
Same classifier in Mandarin:	3.701754	2.923246
Same classifier in Thai:	3.581871	2.980263

Even a comparison of the reaction times showed no significant differences between the two groups. This is shown in figure 2 where the medians are close together.

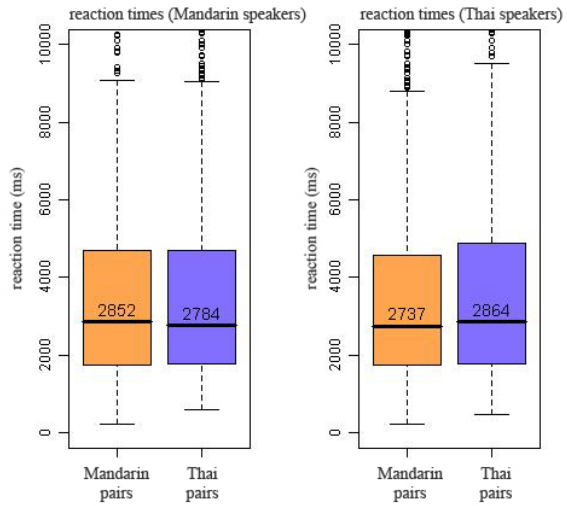


Figure 2: Reaction times

The fact that the participants were not influenced by their classifier systems could perhaps be explained by the exposure of the tested subjects to a non-classifier language, i.e. German. This exposure not only takes place by learning a non-classifier language but during language immersion—a term we borrow from language learning research. The original meaning of language immersion is that children speaking just one language participate in a program in which they "enter a school in which the foreign language is the only medium of instruction" ([24], p. 182). By immersion we mean that the participants were highly exposed to the German language for a long period of time.

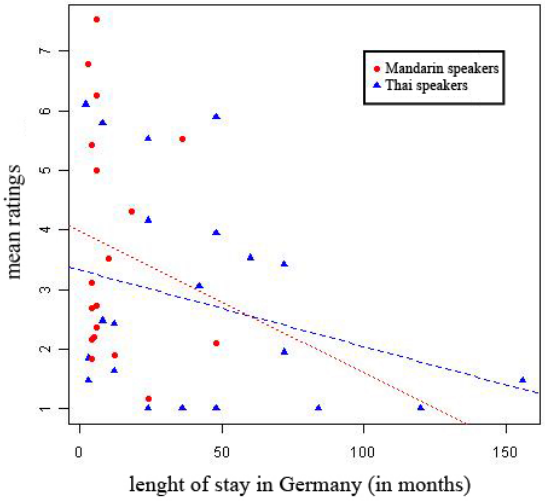


Figure 3: Decreasing mean ratings separated by native language

Therefore, we hypothesized that the longer the participants had been immersed in the German language, the less they would have found similarities in the picture pairs referring to nouns sharing the same classifier in their particular language. Figure 3 separated by native languages, shows that the longer the participants had been living in Germany, the less similar they rated the picture pairs sharing the same classifier in their language. As we see, the average length of stays of the Mandarin speakers are very short compared to the length of stay of the Thai speakers who had a wider range of lengths. Figure 3 makes it clear that there is too little data to draw concise conclusions, so a Pearson's product-moment correlation and a Spearman's rank correlation of the data showed in figure 3 led to a weak negative correlation (-0.1919396 respectively 0.1246900). But the figure suggests that the classifier effects decrease rapidly during the first two years of stay in Germany.

This observed tendency is supported by findings of [5] and [25]. [5] tested the role of shape classifiers in Mandarin. They found that "reliance on shape by Mandarin speakers increases when exposure [...] to the Mandarin language increases." [25] showed that the way a language expresses time has an influence on the speaker's conceptualization of time in non-linguistic tasks (e.g. [26] or [27]). Speakers of Greek, a language which tends to express time in terms of an amount of a substance (see [26], p. 71 and [25], p.

129), show significantly different results in non-linguistic tasks than speakers of English—a language which tends to express time in terms of distance. [25] tested whether these results could be inverted by training English speakers to speak of time in terms of an amount of a substance. In his training experiments he showed "that teaching English speakers to use amount metaphors for time in the laboratory caused them to perform [...] indistinguishably from Greek speakers" ([25], p. 129).
These results are consistent with recent approaches like the one of Australian-American linguist Nicholas Evans who believes that "many of the concepts we use to apprehend the world are built up in the very process of learning to speak – with the result that our conceptual stock differs markedly with our language background." ([31], p. 159). But to test if language immersion weakens Whorf effects or if learning a new language leads to Whorf effects, as the Czech proverb quoted at the beginning suggests, further research should be conducted with more data.

Acknowledgements

We would like to thank Dietmar Zaefferer (University of Munich) who gave rise to the idea of this study and who was very helpful with its design. In addition, Jingyang Xue for her help with Mandarin Chinese; Kornshulee Nikitsch (University of Munich); Krisdi Chairatana (THAI ASA ev.); Christian Körner for their help with Thai; Daniel Casasanto (Max Planck Institute for Psycholinguistics) for his helpful methodological proposals; the Institute of Statistics of the Ludwig Maximilian University of Munich, especially Monia Mahling, Hongbin Xiang, Ariane Straub, Minh Ngoc Nguyen and Carola Kobayashi for their support in statistical questions. We also want to acknowledge to Patric Bach (Bangor University) for his statistical support. Thanks also to Uwe Reichel, Raphael Winkelmann (both University of Munich), Anna Rühl and Jessica Siddins. We want to thank the Department II of the Faculty of Language and Literature of the University of Munich for financial support. In addition we want to thank two anonymous reviewers for their suggestions.

References

[1] J. A. Lucy (1992). *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis*. New York.
[2] M. Imai & D. Gentner (1997). A cross linguistic study on early word meaning: Universal ontology and linguistic influence. In: *Cognition*, 62, 169-200.
[3] S. Zhang & B. Schmitt. (1998): Language-dependent classification: The mental representation of classifiers in cognition, memory, and ad Evaluations. In: *Journal of Experimental Psychology: Applied*, 4(4), 375-358.
[4] B. Schmitt & S. Zhang (1998). Language structure and categorization. A study of classifiers in consumer cognition, judgment, and choice. In: *Journal of Consumer Research*, 25, 108-122.
[5] J. Y. Kou & M. D. Sera (2007). Classifier effect on human categorization: the role of shape classifiers in Chinese Chinese. In: *Journal of East Asian Linguistics*, 18, 1-19.
[6] H. Saalbach & M. Imai (2007). Scope of Linguistic Influence: Does a Classifier System Alter Object Concepts? In: *Journal of Experimental Psychology*, 136(3). 485-501.
[7] M. Y. Gao & B. C. Malt (2009). Mental representation and cognitive consequences of Chinese individual classifiers. In: *Language and Cognitive Processes*, 24(7/8), 1124-1179.
[8] B. L. Whorf (2003). *Sprache, Denken, Wirklichkeit – Beiträge zur Metalinguistik und Sprachphilosophie*. Herausgegeben von Peter Krausser. Hamburg. (First published 1956 as *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*.)
[9] F. Boas (1911). *Handbook of American Indian Languages*. Washington.
[10] I. Werlen (1989). Sprache, Mensch und Welt – Geschichte und Bedeutung des Prinzips der sprachlichen Relativität. Darmstadt.
[11] E. Sapir (1961). *Die Sprache – Eine Einf¸hrung in das Wesen der Sprache*. Translated by Conrad P. Homberger. München. (First published 1921 as *Language: An introduction to the study of speech*.)
[12] R. W. Brown (1976). Reference. In memorial tribute to Eric Lenneberg. In: *Cognition* 4, 125-153.
[13] E. H. Lenneberg (1953). Cognition in Ethnolinguistics. In: *Language* 29, 463-471.
[14] K. Allan (1977). Classifiers. In: *Language*, 53, 2, 285-311.
[15] H. Zhang (2007). Numeral classifiers in Mandarin Chinese. In: *Journal of East Asian Linguistics*, 16, 43ñ59.
[16] Y. R. Chao (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
[17] K. L. Adams& N. F. Conklin (1973). Toward a theory of natural classification. In: *Papers from Ninth Regional Meeting Chicago Linguistic Society*, April 13-15. IL: Chicago Linguistic Society, 1-10.
[18] P. Friedrich (1970). Shape in Grammar. In: *Language*, 46, 379-407.
[19] D. Inglis (2003). Conceptual structure of numeral classifiers in Thai. In: E. H. Casad and G. B. Palmer (ed.): *Cognitive Linguistics and Non-Indo-European Languages*. Berlin: Mouton de Gruyter.
[20] G. Lakoff (1987): *Women, Fire, and Dangerous Things. What Categories reveal about the Mind*. Chicago & London: The University of Chicago Press
[21] G. Lakoff (1986). Classifiers as a Reflection of Mind. In: C. Craig (ed.): *Noun classes and categorization : proceedings of a symposium on categorization and noun classification, Eugene, Oregon, October 1983*. Amsterdam & Philadelphia: John Benjamins Publishing, 13-52.
[22] C. N. Li & S. A. Thompson (1981). *Mandarin Chinese: a Functional Reference Grammar* Berkeley: University of California Press.
[23] Cohen, J., Flatt, M., MacWhinney, B. & Provost, J. (2008): *PsyScope X Build 53* [Computer program] retrieved from:

http://psy.ck.sissa.it/.
[24] Crystal, D. (1992): An Encyclopedic Dictionary of Language and Languages. Cambridge.
[25] Casasanto, D. (2009). When is a Linguistic Metaphor a Conceptual Metaphor? In: V. Evans & S. Pourcel (Eds.): New Directions in Cognitive Linguistics. Amsterdam: John Benjamins. p. 127-145.
[26] Casasanto, D. (2008): Who's afraid of the Big Bad Whorf? Cross-linguistic differences in temporal language and thought. In: Language Learning, 58(1), 63-79.
[27] Casasanto, D. & Boroditsky, L. (2008). Time in the Mind: Using space to think about time. In: Cognition, 106. p. 579-593.
[28] Sapir, E. (1921). *Language: An introduction to the Study of Speech*. New York: Hartcourt.
[29] Matsumoto, Y. (1993). Japanese numeral classifiers: a study of semantic categories and lexical organization. In: *Linguistics*, 31, p. 667-713.
[30] Aikhenvald, A. Y. (2003). *Classifiers. A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.
[31] Evans, N. (2010): *Dying Words. Endangered Languages and What They Have to Tell Us*. West Sussex: Wiley-Blackwell