Embodiment in first and second language processing

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Dipl. Psych. Daniela Katharina Ahlberg
aus Würselen

Tübingen
2016

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	04.11.2016
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Barbara Kaup
2. Berichterstatter:	Prof. Dr. Doreen Bryant

Danksagung

Bedanken möchte ich mich an dieser Stelle bei all den Menschen, die mich die letzten Jahre, Monate und Wochen unterstützt haben und ohne die diese Arbeit nicht möglich gewesen wäre.

Zuallererst bedanke ich mich bei meinen beiden Betreuerinnen, Prof. Dr. Barbara Kaup und Prof. Dr. Doreen Bryant. Ich bin unendlich dankbar für ihre professionelle, unterstützende und warmherzige Betreuung. Immer waren sie ansprechbar, haben mich aufgebaut und mich auch fachlich weitergebracht. Ich konnte von ihrem Wissen auf dem Gebiet der Psychologie und der Linguistik wunderbar profitieren und habe unheimlich viel gelernt in den letzten Jahren.

Außerdem gilt mein Dank der Graduiertenschule LEAD an der Universität Tübingen, die diese Dissertation finanziert hat. Im Rahmen der Graduiertenschule habe ich viele konstruktive Rückmeldungen zu meinen Projekten und viel Unterstützung auch von meinen Kollegen erhalten.

Besonders bedanke ich mich bei meinen Kollegen von der Abteilung Kognition und Sprache im Psychologischen Institut. Sie hatten immer ein offenes Ohr und standen mit Rat und Tat zur Seite, wenn ich mal wieder ganz aufgeregt ankam, weil irgendwas nicht funktionierte oder auf einmal doch. An dieser Stelle bedanke ich mich ganz besonders bei Jessica Strozyk, die mich durch alle Höhen und Tiefen der Datenanalyse und der Reviewprozesse begleitet hat. Danke für deine aufmunternden Worte und, dass du teilweise genauso verzweifelt vor den Daten gesessen hast wie ich.

Vielen Dank auch an Heike Bischoff, danke dass ich mit dir zusammenarbeiten durfte und dich bei den Datenerhebungen immer an meiner Seite hatte. Ich vermisse unsere gemeinsamen Schultestungen schon sehr.

Des Weiteren möchte ich mich bei meinen Freunden bedanken, die Teile dieser Arbeit Korrektur gelesen haben und mich durch Mutmachsätze und positiven Aktivitätenaufbau immer wieder motiviert und unterstützt haben. Ganz besonderen Dank an: Michèle Suhlmann, Katharina Braungart, Rebecca Erschens, Vita Cardinale, Annina Gonzalez, Danina Dreßler und Svenja Köster.

Zu guter Letzt, vielen Dank an meine Familie, dass ihr immer an mich geglaubt habt und an meine bessere Hälfte, danke dass du immer für mich da bist.

Zusammenfassung

Lange Zeit wurde das Feld der kognitiven Psychologie dominiert von amodalen Theorien der Kognition. Diese nahmen an, dass unser Gehirn in verschiedene, voneinander unabhängig arbeitende, Module unterteilt werden kann (Fodor, 1983). Daraus folgend wurde angenommen, dass sich das kognitive System durch abstrakte Informationsverarbeitung auszeichnet. Entsprechend wurden Wahrnehmung und Handlung nur als Input- und Outputmodule betrachtet, unabhängig von höheren kognitiven Funktionen, wie etwa der Sprachverarbeitung. In den letzten Jahrzehnten entwickelten sich wiederum weitere Kognitionstheorien, die unter dem Begriff der verkörperten Kognition zusammengefasst werden und in einem starken Kontrast zu den amodalen Theorien stehen. Diese Theorien der verkörperten Kognition haben die gemeinsame Prämisse, dass das kognitive System auf unserem Körper und dessen Interaktionen mit der Umwelt basiert. Ein bekannter Ansatz innerhalb dieser Theorien ist der Ansatz der Erfahrungsspuren (Zwaan & Madden, 2005). Laut dieser Theorie basiert die Sprachverarbeitung auf der Reaktivierung von Erfahrungsspuren, die von Erfahrungen mit den zugehörigen Objekten, Zuständen oder Ereignissen stammen. Trotz der vielfältigen Belege für diesen Ansatz, die häufig auf handlungsorientierten Kompatibilitätseffekten beruhen, bleiben immer noch einige Fragen offen, die bisher durch die Forschung noch nicht beantwortet werden konnten. Einige dieser Fragestellungen sind Gegenstand der vorliegenden Dissertation. Da viele der Studien bezüglich der verkörperten Kognition Sätze untersuchten, ist immer noch weitgehend ungeklärt, ob die Effekte auf der Verarbeitung einzelner Wörtern innerhalb des Satzes beruhen oder auf der Verarbeitung des Satzes als Ganzes. Daher liegt ein Fokus der vorliegenden Arbeit auf der Untersuchung von Worteffekten. Ein weiterer offener Punkt ist, ob die einzelnen Wortklassen unterschiedlich verarbeitet werden. Daher wurden in Studie 1 Nomen und Verben gegenübergestellt, während in den Studien 2 und 3 Präpositionen untersucht wurden. Die Ergebnisse legen nahe, dass die Information, die in Nomen und Präpositionen enkodiert werden, automatisch verarbeitet werden können, während die enkodierte Information in Verben nur zugänglich ist, wenn die Aufgabe eine bewusste Verarbeitung erfordert. Während sich dieser Teil der Arbeit mit der Verarbeitung der Erstsprache (L1) befasst, ist ein weiteres Ziel der vorliegenden Arbeit, herauszufinden, ob die Theorie der Erfahrungsspuren auch auf das Verarbeiten einer Zweitsprache (L2) anwendbar ist. Hierzu wurden in den Studien 2 und 3 Erwachsene und Schulkinder mit Deutsch als L1 und L2 miteinander verglichen. Die Ergebnisse liefern Evidenz für die Reaktivierung von

Erfahrungsspuren in der Verarbeitung von sowohl der L1, als auch der L2. Des Weiteren legen die Befunde nahe, dass Personen, die Deutsch schon früh als Zweitsprache erlernen, diese in ähnlicher Weise erwerben, wie sie ihre Erstsprache erworben haben. Personen, die erst spät Deutsch als Zweitsprache lernen, zeigen hingegen deutliche Unterschiede und insbesondere einen starken Einfluss der L1 auf die L2. Zusammengefasst sprechen die Ergebnisse dieses Dissertationsprojektes für die Reaktivierung spezifischer Erfahrungsspuren in verschiedenen Wortklassen, wie z.B. Nomen, Verben und Präpositionen. Zudem konnte gezeigt werden, dass Erfahrungsspuren sowohl in der L1 als auch in der L2 reaktiviert werden.

Abstract

For a long time, the field of cognitive psychology was dominated by amodal theories of cognition, which assume that our mind is organized into different modules that work independently of each other (Fodor, 1983). In line with this account, cognition was assumed to be an independent and abstract information processing system, with perception and motor activity working as input and output modules and the language system as its own independent module. However, in the last few decades, a counterpart to amodal theories of cognitive processing has been developed: theories of embodied cognition. The central assumption of these theories of embodied cognition is the premise that cognition is based in the body and its experience with the environment. One well-known and rather specific account within the embodiment framework is the theory of experiential traces (Zwaan & Madden, 2005). According to this account, language comprehension is based on the reactivation of experiential traces that stem from experiencing the corresponding objects, states, or events. Evidence for this account typically derives from action-related compatibility effects. Although there is ample evidence for this account, some issues remain. A few of these problems are addressed in the present dissertation. For instance, although studies on sentence comprehension have provided support for the experiential traces account, it still remains unclear whether these effects can be ascribed to single word effects within these sentences or to the processing of the complete sentences. Therefore, one aim of this dissertation project was to further investigate single word processing. Similarly, it is still unclear whether embodiment effects differ for different word classes. This issue was addressed in Study 1 with a focus on comparing the processing of nouns and verbs. As a result, the embodiment effect for nouns seemed to be rather automatic and task independent, while the processing of verbs provided embodiment effects only in a task in which active processing of the verbs was required. In Studies 2 and 3, the investigation of embodiment effects in single-word processing was extended to the word class of prepositions. The reactivation of experiential traces was shown for different prepositions, but we also found differences among them. Another field with rather sparse evidence for embodiment is the field of second language processing. As of yet, it remains an open question whether the embodiment account can also be applied to second language processing. Therefore, Study 2 addresses first language (L1) and second language (L2) processing in adults, while Study 3 addresses L1 and L2 processing in schoolchildren. The results provide evidence for the reactivation of experiential traces in L1 and L2 processing. In addition, we found differences between early and late L2 learners,

suggesting that early L2 learners are able to acquire the L2 in a similar way as the L1, while the influence of the L1 plays an important role only in late learning of an L2. In sum, the evidence found in this dissertation project supports the experiential traces theory. The language motor compatibility effects obtained in this dissertation project suggest that experiential traces are activated (1) in different word classes as nouns, verbs, as well as spatial prepositions and (2) in L1 and L2 processing.

Table of Content

1 Introduction

The ability to communicate is often considered a critical survival skill. We communicate with each other to express warnings, but also needs and desires. We can listen to each other's problems and provide support, and also receive support in return. By communicating with each other, we make it possible to give others a small insight into our own self, our feelings, and thoughts, things which would otherwise be hidden. But communicating is not always easy, as not all people in the world speak the same language. Around 7000 different languages are spoken across the globe, with some countries such as Papua New Guinea – a country with only 6.5 million citizens (Papua Neuguinea, 2016) -- even home to more than 800 languages (Paul, Simons, & Fennig, 2016). Therefore, when travelling around the world, we often encounter situations in which talking is not always easy, as neither we speak the inhabitants' language nor do they understand our own native language. This often leads to situations in which the body is used to communicate. By pointing or making gestures to support speech, it is still possible to be understood. In colloquial German, this behaviour is referred to with the expression "talking with one's hands and feet". While sufficient communication via body language is no doubt possible, whether the body might also play an important role in general language processing and comprehension is current the subject of much debate. This issue has been addressed in different theories on cognition that are seen as rather contradictory. In the current debate, two main accounts can be distinguished that will be discussed in the present work, the amodal view and the embodied cognition account. In the next section, I will explain both accounts in more detail.

1.1 The amodal vs. the embodiment account of language comprehension

For a long time, the brain was seen as being split into different modules, with every module having a very distinct and specialized role (Dronkers, Pinker & Damasio, 2000). This view drew support, for instance, from research on patients with brain injures, which indicated that two different brain areas, known as the Broca and the Wernicke areas, are responsible for language production and language comprehension, (Dronkers et al., 2000). In accordance with the view that the brain is organized into different modules, theories were developed that also viewed the mind as having a modular structure: the theory of propositional, amodal cognition (Anderson, 1983; Fodor, 1975; Pylyshyn, 1973). The amodal theory of cognition assumes the different modules working relatively independently of another, which means that they are abstract and domain specific. According to this theory, one of these modules specializes in language. Although perception is needed to perceive speech and action is needed in order to form the actual words with one's mouth, this language module is considered to function mostly independently of action and perception (Fodor, 1983; Masson, 2015).

Cognition within the amodal theories is perceived as abstract information and symbol processing. Therein, perception and motor actions are only used as input and output modules (Barsalou, 1999, 2003). For instance, according to a theory on text comprehension within the amodal account (Kintsch & van Dijk, 1978; Kintsch 1988), comprehension of the semantic content of sentences relies on amodal propositional representations. These propositional representations, which are based on the meaning of words and the syntactic structure of a clause, consist of concepts, mental representations of the denoted entities, which are stored in an individual's semantic memory (Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1975; Martinez de la Vega, 2013). A proposition always consists of a set of predicates and arguments, which are represented in the form of predicate calculus. For a sentence like *Emma kicks the ball,* this propositional representation is as follows:

Kick (Emma, ball)

The theory of amodal cognition has gained wide acceptance over the years. However, as research in the field of language has expanded, not only supporting evidence (e.g., Goetz, Anderson, & Schallert, 1981; Kintsch & Keenan, 1973; Ratcliff & McKoon, 1978; Snodgrass 1984; Theios & Amhrein 1989), but also results contradicting the theory have been obtained (Glaser, 1992; Glenberg, Schroeder, & Robertson, 1998; for a review see Barsalou 1999).

One conceptional issue that amodal theories encounter is known as the grounding problem (Harnad, 1990): Propositions in amodal theories are seen as cognitive representations in the form of symbolic codes, with symbols that are amodal and arbitrary (Barsalou, 1999). Symbols are therefore not grounded, making it unclear how they receive meaning (Glenberg, 1997; Martinez de la Vega, 2013).

Another contradiction arose from studies using brain-imaging techniques such as functional magnetic resonance imaging technique (fMRI) and electroencephalography (EEG). In these neuropsychological studies, no single distinct module in our brain was found to be responsible for language processing, but rather several different areas. For instance, several different regions, including the sensory and motor cortex, are activated in the processing of action verbs, and this activation pattern resembles the activation pattern arising when the described action is executed (e.g., Pulvermüller, Härle, & Hummel, 2001). In addition, behavioural studies have found interactions between language processing and motoric actions (e.g., Glenberg & Kaschak, 2002; Olmstead, Viswathan, Aicher, & Fowler, 2009). These results cannot be explained by amodal theories, which view the motor system as working independently of the language system.

As a result, new accounts have been developed that are able to explain the above results by postulating one modal system. These accounts are generally subsumed under the term *embodied cognition*. Other synonyms are *grounded cognition* or *embodiment*. The idea is not new, as Lotze (1852), James (1890), and even Aristotle (4th century BC/1961), and Epicurus (4th century BC/1994) postulated early forerunners of the embodied cognition account. However, no single theory, but rather different theories and accounts with different emphases, exist within the embodiment account (Wilson, 2002). Despite their different emphases, all of these theories share the main assumption that the mind is not independent of the body. Cognition is seen as being dependent on the physical interactions of the body with the environment, as the body experiences and perceives the world surrounding us. This leads to the assumption that an understanding of the mind is only against the backdrop of the relationship between one's own body and the environment. Sensory and motor perception has an influence on central cognition processes and vice versa. This is in contrast to the purely executive role of the body within the amodal framework. The embodiment view leads to the assumption that similar mechanisms and areas are activated in both central cognition processes and direct interactions or experiences with our environment. Accordingly, the meaning of linguistic stimuli is represented not only in language areas, but also directly and automatically in the sensory and motor systems (Pecher & Zwaan, 2005).

One of the first attempts to provide an underlying mechanism was Barsalou's theory of perceptual symbol systems (1999, 2003). According to his theory, a number of associated brain areas become activated while perceiving our environment. The different aspects of an object are assessed multimodally through the different channels of the sensory system. After that, these are represented through neural activation in the associated areas. As part of this, different characteristics of a specific experience, such as form, colour, actions, noise, smell, movements, or associated emotions, are represented. With the help of selective attention, these properties, or our experiences with them, can be saved in long-term memory and serve as symbols later on. In contrast to symbols in amodal theories, the symbols here are modal, as they are saved by type in associated areas. For instance, noises are saved in the auditory domain, temperature and textures in the haptic domain, etc. According to Barsalou (1999, 2003), these are then organized in a kind of simulator, which helps to generate endless different simulations of an object or an event, even if the object is not present or the event is not happening right now. It should also be possible to recombine, integrate, or extend different simulators to develop new, more complex simulations, such as new situations that never happened exactly in that way. Whenever we encounter a known object, whether we see it, read about it, or hear about it, a simulation is activated and thereby facilitates our cognition. These simulations are not seen as a conscious complex mental imagination. Rather, these simulations occur unconsciously in nature and are understood as an automatic process.

One account that builds on Barsalou's perceptual symbol systems theory (1999) is Zwaan and Madden's (2005) theory of experiential traces. This theory states that every interaction with the environment leaves 'experiential traces' in our brain (Zwaan & Madden, 2005). Every time information referring to these interactions is accessed, these experiential traces can become reactivated. Thus, mental representations are based on our experiences, as they were built by connecting experiential traces. Two main sorts of mental representation can be identified according to this theory: reference representations and linguistic representations. While reference representations can be built through perception of and interaction with our environment, linguistic representations can be built whenever linguistic information is received or produced, such as by hearing or speaking. These different representations can be connected in an experiential trace by way of co-occurrences (Hebb, 1949, cited by Zwaan & Madden 2005). According to Zwaan & Madden (2005), examples of such co-occurrences include ducks swimming in a pond or lake, monitors standing on a desk, clouds that are always located in the sky, or branches that are always situated above roots. On the basis of these spatial and temporal co-occurrences, different combinations of objects, events, actions and bodily states can be part of the same experiential trace. Moreover, the experiences connected in an experiential trace can be derived from different modalities. Thus, the experiential traces and their mental representations are multimodal. Applied to language learning, this means that since linguistic labels often co-occur with the objects, situations, and events to which they refer, the corresponding experiential traces eventually become associated with one another. As a result, during later processing of the linguistic label in isolation, the associated experiential traces stemming from interacting with the referents of the linguistic labels become re-activated. In its strongest version, this account of language comprehension thus suggests that comprehension is based on the reactivation of experiential traces. For instance, when we hear or read the word *football*, various experiential traces associated with footballs become re-activated, including traces stemming from seeing a football, feeling a football, kicking a football, listening to a football match, etc. All of these re-activated traces together make up the meaning of the word und thus promote comprehension.

In the following section, I will further describe the empirical evidence that supports the embodiment account by focusing on neuropsychological and behavioural studies.

1.2 Neuropsychological studies supporting the embodiment account

As stated above, a typical example of findings from neuropsychological studies that contradict amodal theories and support embodiment theories is the finding that hearing or reading a sentence activates different areas in our brain in a very specific manner, similarly to the execution of the described actions. In addition to the study by Pulvermüller et al. (2001), Hauk, Johnsrude, and Pulvermüller (2004) reported similar effects in an fMRI study. They first assessed the brain areas activated while performing finger, feet and tongue movements and then compared these activation patterns with those during a passive reading task of face-, foot- and arm-related action words (e.g., to *lick*, to *pick*, to *kick*). The study revealed clear effector-specific activation in the premotor and primary motor areas during language processing. This activation was similar to the conditions in which the participants performed the corresponding actions and was measurable in addition to the activation of areas typically involved in semantic processing. While this study supported the embodiment framework in general, it also showed that simply perceiving the stimuli without performing any additional tasks is already sufficient to evoke these activation patterns; no deeper processing is needed. Similar somatotopic activation was also reported in several other neuropsychological studies, for instance using fMRI while focusing on single action verbs (e.g., Rüschemeyer, Brass, & Friederici, 2007; Kemmerer, Castillo, Talavage, Patterson, & Wiley, 2008; Willems, Toni, Hagoort, & Casasanto, 2010), or phrases or sentences (e.g., Aziz-Zadeh, Wilson, Rizzolatti, & Iacoboni, 2006; Boulenger, Hauk, & Pulvermüller, 2009; Tettamanti et al., 2005), and also between phrase reading and action observation alone (Aziz-Zadeh et al., 2006). Likewise, Klepp et al. (2014) compared activations found during passive reading of hand and foot-related action verbs with motor field sources for actual hand and foot movements in a study using magnetoencephalography (MEG). Klepp et al. (2014) found larger hand source peak amplitudes for hand than foot words and marginally larger foot source peak amplitudes for foot than hand words, respectively. Lastly, studies involving transcranial magnetic stimulation (TMS) also support the idea of motor involvement in language processing (Buccino et al., 2005; Boulenger et al., 2008; Pulvermüller, 2005).

Although some questions about the precise location and functional overlap of motor and language functions exist (e.g., Postle, McMahon, Ashton, Meredith, & Zubicaray, 2008; Klepp et al., 2014), the aforementioned findings provide evidence for an early, and likely automatic, involvement of the motor system in the processing of action-related language (Vigliocco, Vinson, Druks, Barbar, & Cappa, 2011).

In addition to neurophysiological studies, behavioural studies have also been conducted to investigate the nature of these effects. In this way, it is also possible to assess whether the similarity in the activation patterns is just a side effect or whether it also has direct implications for language processing.

1.3 Empirical evidence in behavioural studies on sentence processing

In behavioural research, observed interactions between language and visual processing or motor processing are typically taken as strong evidence for an embodied model of language comprehension (Ahlberg, Dudschig, & Kaup, 2013). For example, Zwaan, Stanfield, and Yaxley (2002) reported that sentence processing could activate very specific visual images. In their study, participants had to process sentences such as "*The girl saw the egg in the frying pan*" and subsequently respond to pictures of the target entity (egg). The pictures could either match the form of the entity described in the sentences (e.g., a fried egg sunny side up) or be in a different form (e.g., an unbroken egg). Responses were faster in the matching than in the mismatching condition, suggesting that readers had a visual representation of an egg in a frying pan available after reading the corresponding sentence.

A good illustration for the reactivation of motor representations during language comprehension is the action-sentence compatibility effect (ACE) first observed by Glenberg and Kaschak (2002). In their study, participants were asked to read sentences and judge the sensibility by moving their arm away from or towards their body. Responses were faster when the movement direction implied by the sentence matched the response movement (e.g., *You opened the drawer* and a movement towards the body) compared to when there was a mismatch (e.g., *You closed the drawer* and a movement towards the body). These results are in line with the assumption that participants reactivated the implied movements when processing the sentences, which then primed the response movements in the matching conditions (Glenberg & Kaschak, 2002). Such compatibility effects between language and motor processing have also been shown in other studies using different kinds of paradigms and materials. Zwaan and Taylor (2006), for instance, conducted a quite similar study and found a compatibility effect for clockwise vs. counter-clockwise rotations. Their participants responded faster to sentences indicating a clockwise rotation (e.g., *Jane started the car*) with the clockwise turning of a knob, compared to the counter-clockwise turning of a knob and vice versa.

While the studies reported thus far focused on a specific movement or movement direction, Scorolli and Borghi (2007) also reported influences of sentence understanding on effector-specific behavioural responses. In this case, the sentences implied the usage of a specific effector (e.g. hand vs. mouth). Participants had to judge the plausibility of sentences with nouns and verbs that refer to objects and actions associated with specific effectors, e.g., *to unwrap* vs. *to suck the sweet*. In the first block, hand and mouth sentences were tested, while hand and foot sentences were tested in the second block. Half of participants had to respond by saying “yes” into a microphone and the other half had to press a foot pedal. They found a compatibility effect between mouth and foot sentences and mouth and foot responses relative to hand sentences.

The studies described above are in line with the embodiment account, as the found compatibility effects suggest that the connected experiential traces become reactivated. However, the question remains as to whether compatibility effects of this type can be ascribed to sentence comprehension or single word comprehension. The original ACE (Glenberg & Kaschak, 2002) was based on sentences, and the same holds true for the studies extending the ACE (e.g., de Vega, Moreno, & Castillo, 2013; de Vega & Urrutia, 2011). However, in one of Zwaan and Taylor’s (2006) experiments, sentences were presented word by word, which revealed a specific compatibility effect on the verb of the sentence (i.e., *opened*). This suggests that the compatibility effect is not due to a sentence wrap-up, but rather depends on the word which defines the action, namely the verb of the sentence. To further investigate this issue, many studies focussing on single word processing have been conducted, which I will provide a short overview of in the next section.

1.4 Current state of research in behavioural studies on single word processing

When it comes to investigating single word effects, nouns referring to concrete entities and verbs referring to concrete actions have been most extensively examined within embodiment research (Vigliocco et al., 2011). However, while these two word classes have been investigated comprehensively in neuropsychological studies (for reviews see Kutas, VanPetten, & Kluender, 2006; Barber and Kutas, 2007), to date only a small number of behavioural studies have investigated differences between nouns and verbs (Vigliocco et al., 2011). Tremendous support for the embodiment account has been found in behavioural studies focusing on the interactions between motor processing and the processing of nouns referring to concrete entities. Nouns can provide information about several aspects and features of the object they refer to. For instance, according to the experiential traces account, if we encounter the word *airplane*, we automatically know the typical shape of the object, certain noises with which it is associated, and the typical location of an airplane, as we most frequently see airplanes in the sky. That we indeed reactivate the location information encoded by a noun referring to an object that typically or exclusively occurs in a certain location has been shown, for instance, in studies focusing on the compatibility between the visual domain and language (Zwaan & Yaxley, 2003; Šetić & Domijan, 2007). Participants were able to recognize words faster if they were displayed in a compatible position on the screen. For instance, the word *root* was recognized faster when it was displayed in the lower part of the screen, while *roof* was recognized faster when displayed in the upper part of the screen. These results show that contextual information regarding, for instance, location is accessed when reading a word.

Furthermore, Lachmair, Dudschig, De Filippis, de la Vega & Kaup (2011) investigated object words by focusing on the compatibility between language and motor responses. In their study task, participants had to respond with either an upward or a downward movement of their hands, and with a subsequent up or down button press to words referring to objects with a typical up or down location (e.g., *root* = down; *cloud* = up). Lachmair et al. (2011) obtained a compatibility effect with faster response times for compatible trials (up word and upwards button press; down word and downward button press) compared to non-compatible trials (up word and downward button press; down word and upward button press). This facilitation effect was found in a lexical decision task, in which conscious processing of the word is needed to fulfil the task, but also in a Stroop-like task (Stroop, 1935). In the Stroop-like task, the words were presented in four different colours that were matched to the two response directions. Thus, participants had to only consciously process the colour of the words to make their response decisions, while word processing was not required. This is seen as support for automatic activation of experiential traces connected to the location information encoded in object words.

Most of the studies reported above found compatibility effects. However, it is important to note that several studies also found interference effects (Estes, Verges, & Barsalou, 2008; Kaschak et al. 2005; Richardson, Spivey, Barsalou, & McRae, 2003). In these studies, faster responses were obtained in non-compatible trials than in compatible trials. Although both kinds of results have been interpreted in line with embodiment theory, and with the experiential traces account in particular, this also reflects a point that has not been fully explored yet. Research suggests that the type of task and the timing might be responsible for the different results. It seems possible that interference occurs in the early stages of word processing, while facilitation effects arise in later stages (e.g., Boulenger et al. 2006; Borghi, 2011; de Vega et al., 2013; Borregine & Kaschak, 2006; Lachmair, et al. 2011; Chersi, Thill, Ziemke, & Borghi, 2010).

The previous findings concentrated on nouns and encoded location information, but it might also be important to look at effector-specific words and whether they activate the corresponding body part, in this case the effector (e.g., hand or foot). This is of special interest as the aforementioned neurophysiological studies concentrated on this sort of words in demonstrating effector-specific activation in the motor cortex. While Scorolli and Borghi (2007) found effector-specific motor activation present in sentence comprehension, up to now, only a few behavioural studies have focused on effector-specific motor activation during word processing, including Marino, Gough, Gallese, Riggio, and Buccino (2011). Marino et al. (2011) investigated the effects of hand-related or foot-related Italian nouns referring to concrete objects (e.g., pencil), and abstract entities (e.g., jealousy) on hand movements in a go-no go paradigm. Participants had to press a response key with their index finger only when presented with concrete objects. Additionally, participants had to wait to respond until an early (150 ms) or late (1150 ms) go signal was delivered after word presentation. The results showed that participants (all right-handed) responded more slowly with their right hand to hand-related words compared to foot-related words. In contrast, with their left hand, they were faster for hand-related words than for foot-related words. Those effects were only found in the early go signal condition. Marino et al. (2011) explained those results with a left hemispheric specialization of language processing. For right hand responses, interference took place due to the left hemisphere being activated by both language processing and motor response activation, which competed for common resources. The authors argued that this kind of interference was not present for left hand responses because the motor activation took place in the right hemisphere and thus did not overlap with activation from language processing. The authors themselves state that this explanation cannot account for the facilitation effect of the left hand, because it does not predict a difference between hand and foot-related words.

Likewise, Mirabella, Iaconelli, Spadacenta, Federico, and Gallese (2012) also used a go no-go paradigm in which arm reaching movements were executed following the presentation of action verbs related to hand or foot actions, while participants refrained from moving when abstract verbs were presented. Mirabella et al. (2012) found greater interference effects on hand movements for hand-related verbs than foot-related verbs, until an SOA of about 500-600 ms. For later SOAs as well as when participants responded only to the colour of the presented words, this effect vanished. Although this result supports the view of early response interference, it remains unclear why no difference between hand and foot-related words was found for late SOAs.

While the studies of Mirabella et al. (2012) and Marino et al. (2011) used only one response effector, namely the hand, Ahlberg et al. (2013) further investigated effector-specific motor activation using both hand and foot as response effectors. Ahlberg et al. (2013) presented four different groups of German words, namely action verbs (e.g., *grasp* vs. *kick*), nouns containing the lexemes *hand* or *foot* (e.g., *handbag* vs. *football*), and nouns referring to objects that are typically manipulated by the effectors (e.g., *cup* vs. *stirrup*), as well as a control group of up/down nouns referring to entities typically located in the upper or lower vertical space (e.g., *roof* vs. *root*) in a modified Stroop paradigm (Stroop, 1935) that had already been employed in earlier studies on nouns referring to special entities (e.g., Lachmair et al., 2011; Dudschig, Lachmair, de la Vega, De Filippis, & Kaup, 2012). Here, participants had to respond to the font colour of the words with a hand or foot button press. The results showed compatibility effects, with shorter response times in compatible trials (e.g., hand response for *cup*; foot response for *stirrup*) than in incompatible trials (e.g., foot response for *cup*; hand response for *stirrup*) for all the three noun groups. However, no differences in response times were obtained for the action verbs. This was surprising in light of Zwaan and Taylor's (2006) results, as well as due to the fact that neuropsychological studies (e.g., Hauk et al., 2004; Pulvermüller et al., 2001) have focused on exactly these verbs to support the embodiment view of language processing. The results indicate that different word groups are processed differently when it comes to activating effector-specific experiential traces. Nevertheless, the reasons behind the processing differences obtained remain unclear (e.g., temporal characteristics of the reading process, characteristics concerning the level of processing, etc.).

Taken together, all reported studies can be seen as evidence for the experiential traces account (Zwaan & Madden, 2005). The obtained results cannot be explained with an amodal view of language comprehension. Evidence comes from neurophysiological studies as well as behavioural studies and is supported by research on sentence as well as single word processing. However, although a lot of evidence supports this view, there are still unanswered questions. For instance, it seems difficult to come up with a consistent explanation regarding the underlying mechanisms causing interference or facilitation. Furthermore, it remains unclear whether experiential traces always become activated automatically during word processing, or whether their activation is task- and/or context-dependent, at least for particular experiential dimensions and/or particular word groups.

In addition, most of the empirical evidence for the embodied view of language processing comes from language comprehension research focusing on adult native speakers (for an overview, see Jirak, Menz, Buccino, Borghi, & Binkofski, 2010). Recently, this account has also gained acceptance in research on first language acquisition (e.g., Glenberg, Brown, & Levin, 2007; Marley, Levin, & Glenberg, 2007). However, as of yet, this account has not played a relevant role in research on second language acquisition.

1.5 Grounded cognition in second language acquisition

The empirical evidence for the embodiment account reported so far has focused on first language comprehension. It remains an open question whether the embodiment theories are applicable to the processing of a second language.

As of yet, only a few studies have investigated embodiment effects in second language learners (e.g., Bergen, Lau, Narayan, Stojanovic, & Wheeler, 2010; De Grauwe, Willems, Rueschemeyer, Lemhöfer, & Schriefers, 2014; Dudschig, de la Vega, & Kaup, 2014). However, whether the theory of experiential traces is transferrable to L2 comprehension is certainly an important question. Current research supports the view that a network of experiential traces is built during first language acquisition, but what happens when people learn and use a second language?

Theories of embodied cognition are not framed in terms of second language processing or acquisition. However, different theories of second language acquisition and processing exist that might also be applied to embodiment (e.g., theories on the semantic representation of two languages). Two examples are the Revised Hierarchical Model (e.g., RHM, Kroll & Stewart, 1994) and the Bilingual Interactive Activation Plus Model (e.g., BIA+, Dijkstra & Van Heuven, 2002). Both models propose that the semantic representations of L1 and L2 are shared across languages. The RHM argues that bilinguals build separate lexicons for the L1 and L2, but the two lexicons are connected to a shared conceptual system that contains the meaning of the words in both L1 and L2 (Brysbaert & Duyck, 2010). In contrast, the BIA+ model states that the two languages L1 and L2 share one bilingual lexicon that is integrated across languages and contains the semantic representations of L1 and L2 (Dijkstra & Van Heuven, 2002). Applied to embodiment theories, embodiment effects should be observed in both L1 and L2 processing due to the shared semantic representations across L1 and L2, although both models make different assumptions about the underlying structure of the lexicon.

Other models, such as the Sense Model (Finkbeiner, Forster, Nicol, & Nakamura, 2004), argue that L2 words are represented in a different way than L1 words (Duyck & Warlop, 2009). Nevertheless, their semantic representations partially overlap as they are conceptualized as a number of distributed semantic senses. It is assumed that L2 words activate fewer semantic senses than their L1 counterparts and that most of the senses associated with L2 words are also associated with the L1, but not vice versa (Finkbeiner, Forster, Nicol, & Nakamura, 2004; Wang & Forster, 2010). This suggests that the semantic representations of L2 words may be less detailed. Applied to embodiment theories, this would lead to the prediction that embodiment effects in L2 speakers are reduced or absent, at least when language proficiency is still rather low or learners have little everyday experience with using the language (De Grauwe et al., 2014).

Neurophysiological studies investigating the above theories on semantic representations of L1 and L2 (Illes et al., 1999; Rüschemeyer, Zysset, & Friederici, 2005) suggest that L1 and L2 speakers use the same cortical network to process language. Furthermore, they indicate that bilinguals access a common semantic system for both languages. This supports the view that semantic representations are shared across L1 and L2 and therefore speaks in favour of the RHM and BIA+ Models.

Interestingly, De Grauwe et al. (2014) investigated embodiment effects with a focus on motor activation in L1 and L2 speakers. Their results provide evidence for embodiment in both L1 and L2 speakers on a neural basis. De Grauwe et al. (2014) conducted an fMRI study in which participants, native Dutch speakers (L1) and German native speakers (L1) who had learned Dutch as a second language (L2), completed a lexical decision task on visually presented Dutch motor and non-motor verbs. Participants had to respond only to pseudowords. Both L2 and L1 speakers showed similar increases in activation in the motor and sensory-motor brain areas for motor verbs as compared to non-motor ones. Although German and Dutch are similar to some extent, this effect was found for both cognates (words with the same meaning that look similar in the two languages, for instance *nemen* in Dutch and *nehmen* in German) and non-cognates, which are words that have the same meaning but do not look similar in both languages (e.g., *goeien* in Dutch, *werfen* in German). This result shows the involvement of the motor cortex during motor word comprehension and thus embodiment of these words in L2 speakers. Again, it supports the assumptions of the RHM and BIA+ models.

Dudschig et al. (2014) found further support for the findings above in their behavioural study of automatic word processing, in a Stroop-like task (Stroop, 1935) adapted from Lachmair et al. (2011). Dudschig et al. (2014) investigated whether L2 words referring to objects typically located in the upper or lower vertical space would automatically activate location-specific experiential traces and thus facilitate upward or downward responses, respectively. Their participants, German native speakers who had learned English as L2, saw L2 English nouns like *star* and *root* presented in different colours and were instructed to respond with an upward or downward movement depending on the font colour. The results revealed a typical compatibility effect, with faster responses for compatible trials compared to incompatible trials. For instance, words such as *star* facilitated upwards responses, while words such as *root* facilitated downwards responses. Moreover, the obtained results were comparable with the results found for L1 words. These findings indicate that location information is not only automatically activated in L1 processing, but also in L2 processing. Thereby, it suggests that the reactivation of experiential traces is not restricted to L1 processing, as it also takes place in L2 processing.

So far, the evidence speaks for the presence of embodiment effects in a second language. But the underlying mechanism or representation is still unclear. In the case described above, categorizations in the target and the source language were similar. Thereby, it is possible to integrate newly-learned words into the already existing network of experiential traces. Inevitably, this leads to the question of what happens to words in a L2 that cannot be linked to a previously built experiential trace. This might occur if the L1 does not have the same categorization or has a totally different categorization. Are L2 words nevertheless integrated into the already existing network of experiential traces of the L1, and if so, how is this accomplished?

An account that might help in answering this question is the Thinking-for-Speaking Hypothesis by Slobin (1996), which states, "each native language has trained its speakers to pay different kinds of attention to events and experiences when talking about them" (1996). He also states that this training occurs during childhood, and accordingly, once a category is set in the L1, it shows strong resistance to post priori restructuring. The L1's categorization is seen as influencing our perception and thereby guiding attention in the L2. This leads to the assumption that perceived categorization in the L2 is directly influenced by the L1 (Lucy, 2011).

Applied to the experiential traces theory, this suggests that the experiential traces built in a first language guide the learning of a second language. In a first step, the words and categorizations of an L2 are added to the L1. For instance, as the learner's attention is shaped by the L1, he or she might first search for similarities between his L1 and L2. Similar words or meanings between the two languages can be linked. However, as the meaning of words or the categorizations in L1 and L2 do not overlap completely, in a second step, already-established experiential traces would need to be restructured. For instance, if cognates between L1 and L2 exist that cover different meanings, experiential traces might need to be restructured. These "false friends" might be connected in an experiential trace according to their appearance, even though the meaning is different. Later, these connections might need to be restructured to match the actual meaning of the words. Furthermore, in the case of different categorizations of similar-looking concepts in L1 and L2, new experiential traces might need to be built. The difficulty of this restructuring process could be reflected in the difficulty of acquiring these differences. A typical example for categorization differences between languages is spatial categorization.

Categorizations of space vary widely across languages (Bowerman, 1996; Stringer, 2010), which leads to certain difficulties in the acquisition of prepositions (e.g., Alonso, Cadierno, & Jarvis, 2016; Bryant, 2012; Coventry & Garrod, 2004; Coventry, Guijarro-Fuentes, & Valdés, 2012; Grießhaber, 1999; Ijaz, 1986; Lütke, 2008; Munnich & Landau, 2010). For instance, with respect to the categorization of the upper subspace, some language make a further differentiation regarding the feature contact, as for example German: An object situated above another object and having contact with it is described using *auf* (*on*). If the described object has no contact with the other object, *über* (*above*, *over*) is used. In contrast, other languages such as Turkish (Becker & Carroll, 1997) do not make this distinction. In Turkish, the word *üstünde* is used in both sorts of configurations (Becker & Carroll, 1997).

An additional question is whether differences in experiential traces can be found in dependence of the age in which the L2 is acquired, as it is well known that the age at which a language is being acquired plays an important role in the acquisition process (Hyltenstam & Abrahamson, 2003; Johnson & Newport, 1989; Meisel, 2009).

In late L2 learning, for instance when the L2 is acquired after the age of six, also referred to as childlike L2 acquisition (6-12 years) or L2 acquisition as teenagers and adults (after 12 years) by Klein (1992), the categorization of the L1, as well as the connected experiential traces might be already strongly consolidated. In contrast, in early L2 learners, for instance when the L2 is learned before the age of six, also referred to either simultaneous bilingualism (0-3 years) or early-successive bilingualism (3-5 years) as defined by Rothweiler and Kroffke (2006), this might be more flexible, allowing the L2 to be learned in a similar way as the L1. For instance, research on German language acquisition shows that if acquisition of the L2 takes place within a child’s first two years, the acquisition process is comparable to that of German as L1 (Tracy & Gawlitzek-Maiwald, 2000). Parallels between L1 and L2 acquisition can still be seen in early L2 acquisition, but here already differences arise: While the acquisition of syntax and the case system is similar to L1 acquisition, the acquisition of the genus system and prepositions differ. They seem to be more similar to late L2 acquisition among adolescents and adults (Kaltenbacher & Klages, 2006). It is also important to note that the later the L2 is learned, the greater the differences between L1 and the L2 acquisition of German become. L2 learners experience greater difficulty in acquiring the L2 with increasing age of acquisition (Hyltenstam & Abrahamson, 2003; Long, 1990; Rösch, 2011). This suggests that the age of acquisition might be a moderating factor in the learning of an L2 and thereby in the building of experiential traces. While in early L2 acquisition, experiential traces might be built independently of the experiential traces in the L1, late L2 acquisition first involves a reliance on L1 experiential traces, which later might need to be restructured.

1.6 Aim of the present dissertation project

The main aim of the current dissertation project is to further investigate the experiential traces account with a focus on single word processing. It seems worthwhile to concentrate on this, as the reported evidence regarding single word processing is still rather mixed. Therefore, in Study 1 the processing of nouns referring to certain entities and action words is contrasted by conducting three experiments. This is done to investigate the reasons for the processing differences with regard to activating effector-specific experiential traces, with a focus on temporal characteristics of the reading process as well as characteristics concerning the levels of processing.

Additionally, in Studies 2 and 3, the processing of spatial prepositions is investigated in order to extend the evidence for the reactivation of experiential traces to another word class. The focus of Study 1 lies on first language processing, while in Studies 2 and 3 the embodiment account will also be investigated in light of second language processing. These two studies focus on whether the processing of a second language is embodied and shows signs of the reactivation of experiential traces. Furthermore, I investigate the potential role of categorization differences between first and second languages, with a focus on the processing of German spatial prepositions. In Studies 2 and 3, L2 speakers of German with different L1s were compared with L1 speakers of German. The L2 speakers were categorized according to whether the spatial categorizations in their L1 differentiated the upper subspace in a similar way as German (e.g., English, Russian) or made no distinction (e.g., Korean, Turkish).

Additionally, the age of acquisition factor was investigated between subjects in Studies 2 and 3. In Study 2, late L2 learners who mainly learned the L2 after the age of twelve were investigated, while in Study 3 we studied children and thus early learners of German, who learned German before the age of six. Furthermore, the effect of language proficiency was taken into account, as the restructuring of experiential traces might be dependent on language proficiency in the sense that more proficient L2 speakers might already have successfully restructured experiential traces, in contrast to L2 speakers with lower proficiency.

Taken together, the evidence gathered in this dissertation project aims to further extend the evidence for the experiential traces account. The hope was that the results would also clarify some unresolved issues in first language processing and provide meaningful insights into second language processing.

2 Conducted Studies

Within this section the three conducted studies will be explained in more detail with regard to methods and results, the latter of which will be discussed in the discussion section afterwards.

2.1 Study 1 - Effector-Specific Compatibility Effects in Nouns and Verbs

Reference:

Ahlberg, D.K., Strozyk, J.V., Dudschig, C., & Kaup, B. (2016). *Processing differences of effector-related nouns and verbs: Discussing effector-specific compatibility effects.* Manuscript submitted for publication.

2.1.1 Summary of Study 1

The aim of the first study was to compare the processing of nouns referring to entities and action verbs. As of yet, only a few studies have examined the differences between nouns and verbs directly (Vigliocco et al., 2011), while a large share of behavioural studies have concentrated on the processing of nouns.

To further address this issue, Ahlberg et al. (2013) investigated effector-specific compatibility effects during single word processing. In this study, different kinds of effector-related words were presented in a modified Stroop paradigm (Stroop, 1935). In particular, four different groups of German words were presented, namely action verbs (e.g., *grasp* vs. *kick*), nouns directly related to the effectors and containing the lexemes *hand* or *foot* (e.g., *handbag* vs. *football*), and nouns referring to objects that are typically manipulated by the effectors (e.g., *cup* vs. *stirrup*), as well as a control group of up/down nouns referring to entities typically located in the upper or lower vertical space (e.g., *roof* vs. *root*). Participants were tasked with responding to the font colour of the words with either a hand button or foot pedal press, resulting in compatible (e.g., hand response on roof/grasp/handbag/cup) and incompatible trials (e.g., hand response on roof/kick/football/stirrup). For hand-related words and up words, compatible conditions consisted of trials in which the correct response involved a key press with the hand. For foot-related words and down words, compatible conditions consisted of trials in which the correct response involved the foot pedal on the ground. The results showed compatibility effects, with shorter response times in compatible trials (e.g., hand response for *cup*; foot response for *stirrup*) than in incompatible trials (e.g., foot response for *cup*; hand response for *stirrup*) for all the three noun groups but not for the action verbs (Ahlberg et al., 2013). This was surprising in light of Zwaan and Taylor's (2006) results, and also given that neuropsychological studies (Hauk et al., 2004; Pulvermüller et al., 2001) have focused on exactly these verbs in providing support for the embodiment view of language processing.

The aim of the present study was to investigate why no compatibility effects for action verbs were observed in the previous study. We focus on four different possibilities. First, timing differences may be responsible for the null effect. Specifically, verbs cover a broader meaning than nouns (Gentner, 1981) and thus possibly require more processing effort in comparison to nouns. Maybe the processing of the nouns conflicted with response selection (hand vs. foot) because the meaning of the noun was available before the response could be selected. In contrast, because verbs require more processing effort, response selection may have taken place before the meaning of the verb was becoming available, thus explaining why the meaning of the verb does not affect response selection. To address this possibility, we conducted a more complex task in the current study (Experiment 1). Instead of using two effectors (right hand and right foot), we now used four effectors (right/left hand and right/left foot). This modification should give participants more time to process the words before selecting the response. Thus, verbs in this case may have been processed to such an extent that verb meaning and response selection come into conflict with one another. If this is the case, then we expect to find a compatibility effect also for the action verbs in this experiment.

Second, depth of processing may be responsible for the observed null effect for the action verbs. Possibly action verbs activate effector-specific information only in tasks that require lexical access and thus deeper processing than is required in a Stroop-like task focusing only on font colour. The results of a study by Mirabella et al. (2012) are in line with this hypothesis. Their participants had to respond with a reaching movement of the left or right arm to action verbs and to refrain from moving, when abstract verbs were being shown. The authors found an interference effect with longer response times for hand- vs. foot-verbs. Interestingly, this interference effect disappeared when instead of the semantic task a Stroop-like task was administered. To investigate this possibility, we administered a lexical-decision task in the current study (Experiment 2). If a task requiring lexical access is needed to find compatibility effects for action verbs, we should find a compatibility effect for action verbs in this experiment.

A third reason for why we did not previously find a compatibility effect for action verbs may have to do with the fact that we presented more nouns than verbs in those experiments. Maybe this has led participants to focus on the nouns and to neglect verb processing. In the third experiment of the current study, we therefore only presented action verbs and manipulated the experimental task (Stroop-like vs. lexical-decision task) within participants. If the biased distribution of nouns and verbs in the experimental setup was responsible for the null-effect, then we should find a compatibility effect in both tasks in this experiment, in which only verbs were being presented. If the depth of processing explanation is correct, we should find a compatibility effect only in the lexical decision but not in the Stroop-like task.

Finally, a fourth possibility would be that action verbs are associated with very specific motor plans. Maybe the movement that an action verb refers to is so specific that it does not conflict with or facilitate a simple button or foot pedal press. If so, then we should not find compatibility effects in any of the three experiments in the current study, because response movements never directly match the specific actions that are associated with the respective action verbs.

In Experiment 1, we employed a more difficult Stroop-like task than in the original study, namely one involving four instead of two effectors (right/left hand and right/left foot). As expected, the mean response time in this study increased. However, we still only observed effector-specific compatibility effects for nouns, not for verbs, which speaks against the idea that timing differences are responsible for the different results obtained for nouns and verbs. In Experiment 2, we presented participants with effector-related nouns and verbs but this time in a lexical decision task. In line with our hypothesis and in line with the results of a study by Mirabella et al. (2012), we now found effector-specific compatibility effects for nouns as well as for verbs. In Experiment 3, we directly compared a Stroop-like task with a lexical decision task in which we presented participants only with effector-related action verbs. In line with our predictions, we found an interaction between compatibility and task. Action verbs only showed effector-related compatibility effects when processed in a lexical decision task, not when processed in a Stroop-like task that does not require participants to access their mental lexicon.

Taken together, the results of the three reported experiments suggest that there is a difference between noun and verb processing in the sense that nouns but not verbs automatically activate effector-specific information. For verbs, participants need to be forced to access their mental lexicon before any evidence can be found that they indeed activate effector-related information. I want to further discuss this issue in the general discussion.

In sum, we found clear evidence for effector-specific compatibility effects for single word reading, both for nouns referring to objects that are typically manipulated with either the hand or the foot, as well as for verbs referring to an action that typically involves either the hand or the foot. As such, the current study provides evidence for the embodied cognition account, which assumes that readers activate experiential traces when reading words or sentences that stem from prior interactions with the referents of the linguistic expressions. Future studies are necessary to determine which differences between nouns and verbs best explain the observed differences in task dependency.

2.1.2 Detailed description of methods and results of Study 1

This section provides details regarding the applied methods and observed results for the three experiments conducted in Study 1.

Methods Experiment 1: Stroop-Like Task with Four Effectors

In the first experiment, participants were presented with effector-related nouns and verbs written in one of four different colours as in the study by Ahlberg et al. (2013). Each colour was mapped to one of four effectors in a Stroop-like task: Correctly responding to the words required a hand or a foot button press on the left or right side, depending on the font colour of the stimuli. If reading an effector-related word activates the respective effector and thus primes responses with this effector or hinders responses with a different effector, then a compatibility effect should be observed in this experiment. More specifically, in case action verbs need more time than nouns to be processed before the respective effector is being activated, and if this is the reason why there was no compatibility effect in the previous study for action verbs, then there should now be a good chance to find such a compatibility effect in the current setup. The reason is that response selection in this task with four effectors is more complex and thus leaves more time for word processing prior to response selection.

Participants. Forty-eight German native speakers, aged 18 to 26 years (7 male; M_{age} = 20.5 years, SD_{age} = 1.7 years) participated for course credit or financial reimbursement after signing a form of consent. The participants had normal or corrected-to-normal vision. We assessed handedness using a translated version of the Edinburgh inventory (Oldfield, 1971). Forty-seven participants were classified as right-handed (M = 81.6; score range: +46.6 to +100), one participant was classified as left-handed (-62.5).

Materials and apparatus. We used the same stimuli as Ahlberg et al. (2013), namely 192 German nouns and verbs, subdivided into four different categories. The first group consisted of 64 hand- or foot-related action verbs (e.g., *grasp* vs. *kick*) originally taken from Pulvermüller et al. (2001). The second group (explicit nouns) consisted of 32 hand or foot related nouns including the lexemes *hand* or *foot* (e.g., *handbag* vs. *football*). The third group (associated nouns) consisted of 32 hand- or foot-related nouns without the lexemes *hand* or *foot* that referred to objects that are typically manipulated with the hand or the foot (e.g., *cup* vs. *stirrup*). The fourth group consisted of a shortened set, namely 64, of up/down words (e.g., *root* vs. *roof*) from the study of Lachmair et al. (2011). This group was included in the original study by Ahlberg et al. (2013) as a control group[1], and – for the sake of comparability – was included in this study as well. However, as the two sets of words in this group do not systematically differ with respect to effector-specificity, we will not include these words in our analyses but rather treat them as filler items.

Stimuli were presented in the colours blue (rgb 0, 0, 255), orange (rgb 255, 128, 0) brown (rgb 140, 80, 20), and lilac (rgb 150, 0, 255) on a white background in centre position on a CRT-screen in Type Size 12 in Courier New bold. Each colour occurred equally often and the colour assignment to the effectors was counterbalanced across participants, resulting in 24 different experimental versions.

In contrast to the study of Ahlberg et al. (2013), in which the participants stood in front of the computer with a height-adjustable table, in this experiment the participants sat in front of the computer. As can be seen in Figure 1, responses were recorded via four buttons (two for the feet and two for the hands) on two keyboards with a constructed overlay. One of these was placed on the table and the other one was placed on the ground. This setup would not have been possible in a standing position, because it would not have been possible to use both feet for responding. The experiment was programmed with E-Prime® (Psychology Software Tools Inc., http://www.pstnet.com/E-Prime/e-prime.htm).

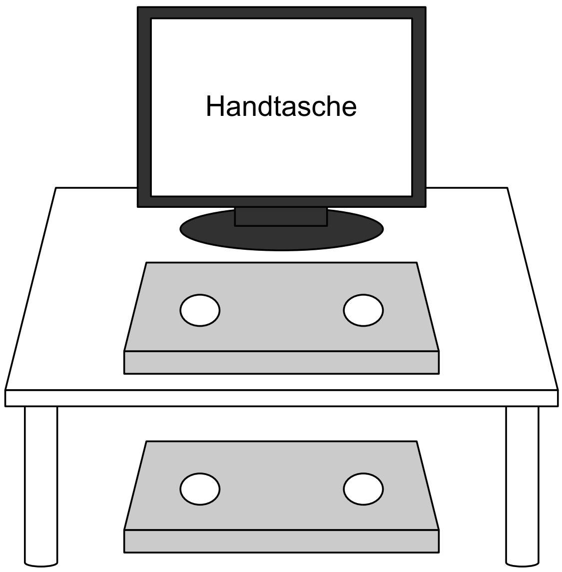


Figure 1. Experimental setup. Two keyboards with a constructed overlay served as response devices. Participants pressed the two buttons on the table with their left and right hand, respectively. The two buttons on the ground recorded responses with their feet. Participants took of their shoes and wore foot covers during the experiment.

[1] Based on the literature, one would expect that up-words are responded to faster with the hand (up-response) than with the foot (down-response) and vice versa for down-words. This was indeed what was found in the study by Ahlberg et al. (2013), and thus allowed to demonstrate the functionality of the experimental setup.

Procedure and design. Each trial started with a fixation cross, displayed in the centre position of the screen for 800 ms. Afterwards the stimulus was presented until the participant responded. Between trials a white screen was shown for 1000 ms.

Every word was presented four times, resulting in a total amount of 768 trials, subdivided into 4 experimental blocks. The experiment started with a practice block, in which 16 stimuli were presented two times each in different colours. These stimuli were not presented in the experimental blocks. In contrast to the experimental blocks, the participants received accuracy feedback during the practice block.

Participants were instructed to respond to the font colour as quickly and accurately as possible. For each participant, each of the four colours was mapped to one effector. The mapping of colours to response directions was balanced across participants: All possible mappings occurred equally often.

The design was a 3 (word group) x 2 (response compatibility) within-subjects design. The dependent variable was the latency of the button press.

Results and Discussion Experiment 1

One participant was excluded from the data analysis due to an error rate above 15%. Mean error rate after exclusion was 4.9%. We excluded error and practice trials. In addition, responses deviating by more than 3 SDs from the mean for each participant and condition (word group x response compatibility) were excluded, which reduced the data by 1.8%. Mean response times of the remaining trials are displayed in Figure 2.

The analyses revealed a significant main effect for compatibility, $F(1, 46) = 10.23$, $p = .003$, $\eta_p^2 = .182$, and a response compatibility-by-word group interaction, $F(2, 92) = 5.25$, $p = .007$, $\eta_p^2 = .102$. There was no main effect of word group, $F(2, 92) = 0.02$, $p = .982$, $\eta_p^2 < .001$.

Separate analyses for the three word groups revealed significant effects for the two effector-related noun groups (explicit nouns: $F(1, 46) = 8.45$, $p = .006$, $\eta_p^2 = .155$; associated nouns: $F(1, 46) = 9.23$, $p = .004$, $\eta_p^2 = .167$) but no significant compatibility effect for the action verbs $F(1, 46) = 0.06$, $p = .809$, $\eta_p^2 = .001$.

This experiment was a replication of the study of Ahlberg et al. (2013), with the main difference being that four instead of two effectors were involved in the experimental task. We replicated the main results of that study, namely finding compatibility effects for the two effector-related noun groups (explicit nouns and associated nouns). Most importantly, the action verbs in the current experiment again did not show a compatibility effect, although the mean response times were 240 ms longer than in the prior study, indicating that we indeed accomplished our goal of making the task more complex. Thus, simply making the response selection more complex to give the language processing system more time to process the verbs before response selection did not lead to a compatibility effect for action verbs. Possibly, action verbs are not processed deeply enough to activate the corresponding effector if the task does not require lexical access. This possibility will be investigated in Experiment 2.

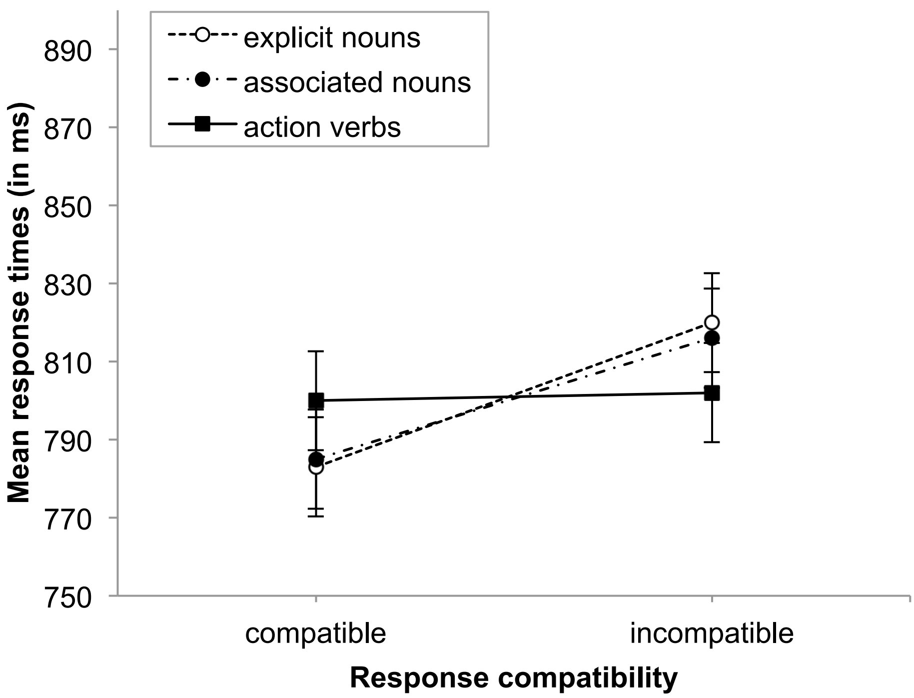


Figure 2. Mean response times of correct responses as a function of response compatibility and word group. Error bars represent 95% confidence intervals (as per Masson & Loftus, 2003).

Method Experiment 2: Lexical-Decision with Two Effectors

In Experiment 2, we wanted to find out whether a compatibility effect would be observed for action verbs if the experimental task required lexical access. Maybe action verb meaning is too complex to be assessed automatically. In the present experiment, we presented participants only with two word groups, namely the explicit nouns and the action verbs. If our hypothesis is correct, and the meaning of action verbs is only processed deeply enough for compatibility effects to occur if the task requires lexical access, then we should now find compatibility effects for both word groups, the explicit nouns as well as the action verbs. In order to keep the conditions as similar as possible to the original experiment by Ahlberg et al. (2013), participants were now responding in a standing position.

Participants. Twenty-four native speakers of German, aged 18 to 44 years (2 male; M_{age} = 23.3 years, SD_{age} = 5.9 years) participated for course credit or financial reimbursement after signing a form of consent. All participants had normal or corrected-to-normal vision. We also assessed handedness using a translated version of the Edinburgh inventory (Oldfield, 1971). All participants were classified as right-handed (*M* = 88.5; score range: +62.5 to +100).

Materials and apparatus. Materials were made up of the explicit nouns (32) and the action verbs (64) from Experiment 1 as well as 96 pseudo words (e.g., zalmen, Hestgeleur). The pseudo words were generated with the help of the pseudo word generator Wuggy on the basis of our stimuli (Keuleers & Brysbaert, 2010). Stimuli were presented in centre position on a CRT-screen in Courier New type size 12 bold.

Responses were recorded via a PST Serial Response Box, Model Number 200A with a foot pedal. The Experiment was programmed with E-Prime® (Psychology Software Tools Inc., www.pstnet.com/E-Prime/e-prime.htm). The participants stood in front of a height-adjustable table, CRT-screen as well as response box situated on it, with the possibility of leaning against the wall with their back. Prior to the experiment, the height of the screen was adjusted such that stimulus words were presented at eye-level of the participants. The foot pedal was adjusted and fixed in a proper distance to the participant. Every participant reacted with his or her dominant side of the body.

Procedure and design. Each trial started with a fixation cross, displayed in centre position of the screen, lasting 800 ms. Then the stimulus was presented until response. Between trials a white screen was shown for 1000 ms.

Participants were asked to perform a lexical-decision task. For half of the participants the response mapping was hand button press in case of “yes” and foot pedal press in case of “no” for the first and the third block of the experiment and the reversed response pattern for the second and fourth block of the experiment. The remaining participants received the reversed order instructions. We measured the response times in this lexical-decision task.

Every word was presented four times, resulting in a total amount of 768 trials, which were subdivided into 4 experimental blocks. At the beginning of each block the instruction changed and therefore each block started with a practice block, in which 22 words (11 words as well as 11 pseudo words) were presented. These stimuli were different from the experimental stimuli. In contrast to the experimental blocks, the participants received feedback about response accuracy during the practice blocks.

The design was a 2 (word group) x 2 (response compatibility) within-subjects design. The dependent variable was the latency of the button or foot pedal press, respectively.

Results and Discussion Experiment 2

The results were analysed as in Experiment 1. Practice trials, error trials, and pseudo word trials were excluded from further analyses. Mean error rate was 4.8%. Responses deviating by more than 3 SDs from the mean for each participant and condition (word group x response compatibility) were excluded. This reduced the data by 1.8%. Mean response times are displayed in Figure 3.

The analyses revealed a significant main effect for word group, $F(1, 23) = 56.67, p < .001$, $\eta_p^2 = .711$, and response compatibility, $F(1, 23) = 14.31, p < .001, \eta_p^2 = .384$. There was no significant interaction effect, $F(1, 23) = 0.23, p = .635, \eta_p^2 = .010$. The separate analyses for the two word groups revealed significant compatibility effects for both groups (explicit nouns: $F(1, 23) = 16.57, p < .001, \eta_p^2 = .419$; action verbs: $F(1, 23) = 6.34, p = .019, \eta_p^2 = .216$).

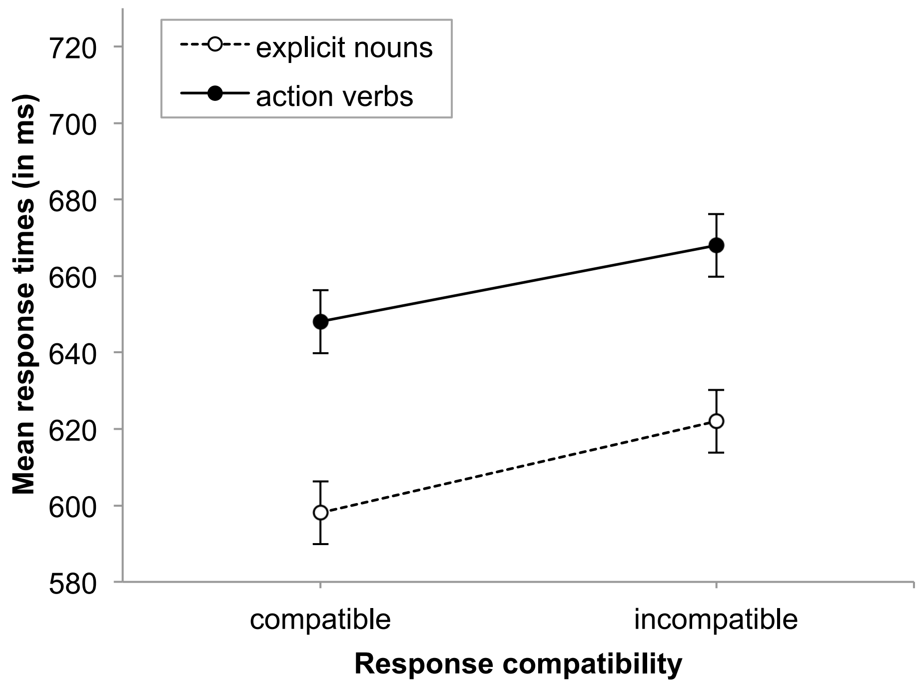


Figure 3. Mean response times of correct responses as a function of response compatibility and word group. Error bars represent 95% confidence intervals (as per Masson & Loftus, 2003).

In this experiment, we now indeed for the first time in our lab found a compatibility effect for action verbs. In our view, the reason most likely has to do with the experimental task administered in this experiment. Whereas in previous experiments our experimental tasks varied in complexity but always focused on superficial properties of the linguistic stimuli (namely font colour), the task in the present experiment required lexical access and thus deeper processing of the presented words. Thus, the results of the present experiment are nicely in line with the hypothesis that effector-related action verbs do not automatically activate the respective effector but only do so when participants are forced to access their mental lexicon. However, before jumping to this conclusion, some alternative explanations need to be ruled out that take into account differences between this experiment and Experiment 1. First, in this experiment, participants saw more verbs than nouns which might have biased participants towards a deeper processing of verbs. Second, in this experiment, participants were standing in front of the computer screen whereas in Experiment 1 they were sitting. Although we do not consider it likely that standing or sitting makes a difference for effector-related compatibility effects, we nevertheless consider it helpful to see whether action verbs also lead to a compatibility effect with participants in a sitting position. In Experiment 3, we therefore presented sitting participants only with action verbs and manipulated task (Stroop-like vs. lexical-decision task) within participants. This allows us to directly investigate the hypothesis that the relevant factor for obtaining an effector-related compatibility effect for action verbs is indeed the experimental task.

Method Experiment 3: Task Manipulated Within Participants

In this experiment, we presented only the action verbs and manipulated the experimental task in a within-subject design. Each participant performed both tasks, half of the participants started with the Stroop-like task and the other half started with the lexical-decision task.

Participants. Forty-eight native speakers of German, aged 18 to 33 (11 male; $M_{age} = 22.9$ years, $SD_{age} = 3.5$ years), participated for course credit or financial reimbursement after signing a form of consent. All participants had normal or corrected-to-normal vision. The handedness of the participants was assessed using a translated version of the Edinburgh inventory (Oldfield, 1971). All participants were classified as right-handed ($M = 80.7$; score range: +50 to +100).

Materials and apparatus. In this experiment, we combined the Stroop-like task and the lexical-decision task in one experiment. We presented only the action verbs of Pulvermüller et al. (2001), the same as in the first two experiments.

In the Stroop-like task, stimuli were presented in the colours blue (rgb 0, 0, 255), orange (rgb 255, 128, 0), brown (rgb 140, 80, 20), and lilac (rgb 150, 0, 255) on a white background, in centre position in type size 12 in Courier New bold. In the lexical-decision task, stimuli were presented in black on white background, centre position in Courier New type size 12 bold.

We used the same setup for both tasks. Stimuli were presented on a CRT-screen and the participants sat in front of the computer. Responses were recorded via a PST Serial Response

Box, Model Number 200A with foot pedal. The experiment was programmed with E-Prime® (Psychology Software Tools Inc., www.pstnet.com/E-Prime/e-prime.htm).

Procedure and design. The two tasks were completed after one another. The order was balanced across participants, half of them started with the Stroop-like task and the other half started with the lexical-decision task.

In the Stroop-like task, the general procedure was the same as in Experiment 1. Every word was presented four times, resulting in a total amount of 256 trials, which were subdivided into 4 experimental blocks. The experiment started with a separate practice block, in which 10 stimuli were presented two times each in different colours.

In the lexical-decision task, the general procedure was the same as in Experiment 2. Here the words were presented twice, distributed over two blocks, resulting in 256 trials in total. The practice block consisted of 10 words and 10 pseudo words and was presented right before the start of each block. For half of the participants the response mapping was hand button press in case of "yes" and foot pedal press in case of "no" for the first block of the experiment and the reversed response pattern for the second block of the experiment. The remaining participants received the reversed order instructions.

The design was a 2 (task) x 2 (response compatibility) within-subjects design. The dependent variable was the latency of the button press.

Results and Discussion Experiment 3

Results were analysed as in the two experiments before. We excluded all error trials and pseudo word trials. Mean error rate was 4.5%. Responses deviating by more than 3 SDs from the mean for each participant and condition (task x response compatibility) were excluded from further analyses. This reduced the data by than 1.7%. Mean response times are displayed in Figure 4.

The analyses revealed significant main effects of task, $F(1, 47) = 60.90$, $p < .001$, $\eta_p^2 = .564$, and response compatibility, $F(1, 47) = 10.14$, $p = .003$, $\eta_p^2 = .177$, as well as a task-by-response compatibility interaction, $F(1, 47) = 5.44$, $p = .024$, $\eta_p^2 = .104$.

The separate analyses for the two tasks revealed a significant compatibility effect for the lexical-decision task, $F(1, 47) = 9.67$, $p = .003$, $\eta_p^2 = .171$, while there was no compatibility effect in the Stroop-like task, $F(1, 47) = 1.26$, $p = .268$, $\eta_p^2 = .026$. These results clearly support the hypothesis that the experimental task is the critical factor for finding an effector-related compatibility effect. In addition, this experiment rules out the idea that effector-related compatibility effects for action verbs are only found in standing position, and also that these effects depend on a material set in which verbs are overrepresented. If the former had been true, we should not have found a compatibility effect in either of the tasks. If the latter had been true, we should have found a compatibility effect in both tasks. The observed interaction of compatibility and task clearly speaks against these possibilities.

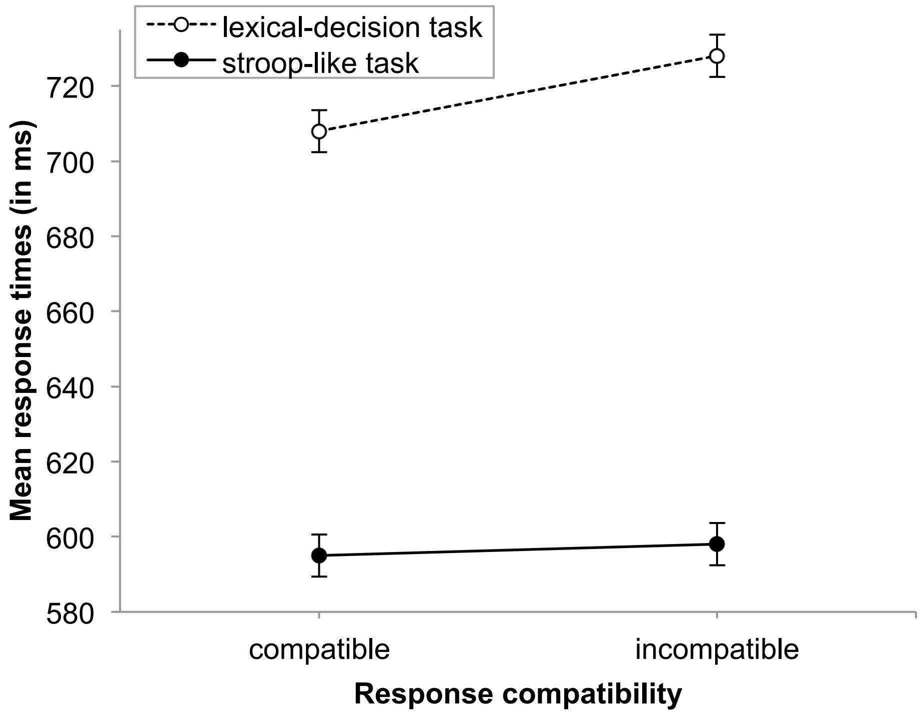


Figure 4. Mean response times of correct responses as a function of response compatibility and conducted task. Error bars represent 95% confidence intervals (as per Masson & Loftus, 2003).

Properties of our Stimulus Sets

An overview of the properties of the different word groups used in our study can be found in Table 1. We included information on word length, word frequency, mean bigram frequency, number of orthographical neighbours, imageability ratings, as well as co-occurrence values with the words *hand* and *foot*. Word class frequencies were retrieved from the "Wortschatz Portal" of the University of Leipzig (http://wortschatz.uni-leipzig.de). Mean bigram frequencies and the number of orthographical neighbours were determined according to Coltheart (Coltheart, Davelaar, Joasson, & Besner, 1977) based on data retrieved from the "dlexDB" corpus (http://dlexdb.de; Heister et al., 2011). Imageability ratings were obtained from Köper and Schulte im Walde (2016). In addition, we determined co-occurence values between each of our stimuli and the words *hand* and *foot*, respectively, based on Latent Semantic Analysis (LSA) (Günther, Dudschig, & Kaup 2015). The connected semantic space *sdewac_hafu* can be found at: http://www.lingexp.uni-tuebingen.de/z2/LSAspaces/.

As can be seen in Table 1, nouns and verbs do indeed differ with respect to a number of variables. Nouns are longer compared to verbs ($t(126) = 4.09$, $p < .001$) and show a higher imageability rating ($t(126) = 2.56$, $p = .012$). Verbs on the other hand are higher in frequency ($t(126) = 4.34$, $p < .001$), higher in bigram frequency ($t(126) = -9.21$, $p < .001$), and they do have more orthographical neighbours than nouns ($t(126) = -4.16$, $p < .001$). No difference between nouns and verbs was observed regarding their co-occurrence with the words *hand* and *foot*, respectively ($t(125) = -0.87$, $p = .385$). We will come back to these differences in the General Discussion.

Table 1: *Comparison of the properties of the different word groups.*

Properties	Action verbs	Explicit nouns	Associated nouns
Length	6.89 (1.20)	8.22 (1.79)	8.03 (2.39)
Word frequency	12.95 (2.85)	14.97 (2.97)	15.31 (2.79)
Mean bigram frequencies	301,075.95 (94,736.96)	142,136.26 (74,860.13)	181,552.45 (71,075.42)
Orthographical neighbours	12.25 (9.47)	4.09 (10.67)	6.16 (9.15)
Imageability	5.74 (1.00)	5.88 (0.85)	6.45 (0.81)
Co-occurrence with *hand/foot* (LSA)	0.50 (0.12)	0.46 (0.23)	0.48 (0.16)

Note. The table contains means with the respective standard deviation in parentheses below.

2.2 Study 2 - Language-space interactions in adult speakers of L1 and L2 German

Reference

Ahlberg, D.K., Bischoff, H., Strozyk, J.V., Bryant, D., & Kaup, B. (2016). *Grounded cognition: Comparing language-space interactions in L1 and L2.* Manuscript submitted for publication.

2.2.1 Summary of Study 2

In the present study, we aimed to investigate how spatial prepositions are processed by learners of German as a second language (L2) as well as by German native speakers who learned German as their first language (L1). We were particularly interested in comparing embodiment effects related to the processing of spatial prepositions in German native speakers with the embodiment effects potentially observed in different groups of L2 learners. For that reason, we investigated German native speakers and compared them to L2 speakers whose native language uses similar spatial terms as German with respect to the upper part of space (English and Russian) and L2 speakers whose native language uses different spatial terms than German (Turkish and Korean). While English uses the prepositions *on* and *above*[2], in Russian the prepositions *на* (*na*) and *над*[3] (*nad*) can be seen as near equivalents to *auf* and *über*. In contrast, Turkish uses either *üstünde* or *üzerinde*[4], which are used interchangeably, and Korean uses only *위*[5] (*wi*) for both spatial configurations +/- contact in the upper subspace. One additional difference is that while English and Russian use prepositions, the above terms in Turkish and Korean are handled as postpositions (Becker & Carroll, 1997; Munnich, Landau, & Dosher, 2001).

Based on Bryant's (2012) studies and on Slobin's (1996) statement about the rigidity of a category once it is set, our hypothesis was that newly learned words would be connected to

---

[2] With regard to frequency, *on* occupies rank 17, *above* occupies rank 896. An alternative expression of *above* is *over* with rank 124. This information was retrieved from the Corpus of Contemporary American English (Davies, 2008).

[3] *на* (*na*) occupies frequency rank 4; *над* (*nad*) occupies rank 181 according to Sharoff, Umanskaya, & Wilson (2013).

[4] Although both words are used interchangeably, *üstünde* with rank 647 is less frequent than *üzerinde* with rank 92 (Aksan et al., 2012)

[5] *위*(*wi*) occupies rank 119 in word frequency, although in the above stated literature *wi* is seen as a postposition, it is also referred to as a common noun (National Institute of Korean Language, 2005). An alternative expression would be the verb *Nohta* (English: *put on*; Choi & Hattrup, 2012).

pre-existing experiential traces of the first language, and that the nature of the embodiment effects found in the L2 would, therefore, depend on the specific L1 of the participants. More specifically, in our study we focused on the prepositions *auf* (*on*) and *über* (*above*), as we predicted differences in the processing of these spatial terms in German native speakers and second language learners of German depending on the nature of their L1. We presented these terms in an experimental setup similar to the one used by Lachmair et al. (2011) and Dudschig et al. (2014). For German native speakers, we expected larger embodiment effects for *über* compared to *auf,* because *über* implies a larger distance between theme and relatum on the vertical axis. In addition, *auf* in contrast to *über* is often used non-spatially in German (e.g., *Ich freue mich auf die Party* '*I am looking forward to the party*'; *aufräumen* '*to tidy up*' etc.). For L2 speakers of German, we predicted different results depending on their particular L1. We expected the English and Russian L1 speakers to show a similar processing difference between *auf* and *über* as the German native speakers, because the split of the upper subspace into +/- contact is also present in English and Russian. For the Turkish and Korean L1 speakers, we predicted a different pattern. These speakers should show stronger effects for *auf* in comparison to *über*. In their L1, a category split for the upper subspace into +/- contact is not present, but since *auf* is much more frequent in the German input (see Ruoff, 1990), we expected them to transfer their experiential traces to this term. The degree to which this is the case might also be dependent on age of acquisition or language experience (De Grauwe et al., 2014), which we measured as well.

Our first experiment, in which German native speakers processed German prepositions in a Stroop-like paradigm in which they responded with an upward or downward directed response movement depending on font colour, served as a kind of baseline experiment to which the results obtained with different groups of people learning German as a second language could be compared. We indeed found a significant compatibility effect in this experiment, reflected in an interaction between the meaning of the spatial preposition and the direction of the response movement. Response times were faster in conditions in which the meaning of the preposition was compatible with the response direction (upward movement for *über*; downward movement for *unter*) compared to conditions in which the two were incompatible (upward movement for *über*; downward movement for *unter*). This result clearly shows that compatibility effects can be observed even with very small units such as prepositions and even if the same terms are presented over and over again to the participants.

In Experiment 2, we then tested speakers whose first language has a similar categorical split along the vertical axis as German, namely native speakers of English or Russian. For these speakers, we also found compatibility effects. Interestingly, in our third experiment, we were also able to show the presence of compatibility effects for participants whose native language does not display a categorical split along the vertical axis, namely for native speakers of Turkish or Korean. For these speakers, the spatial terms from their L1 cannot be matched directly to the German system during second language learning. Nevertheless, compatibility effects were observed, suggesting that the reactivation of experiential traces does indeed not only play an important role in first language processing but also in the processing of a second language, even if this is quite dissimilar to the participants' native language. A relevant question now is whether these experiential traces rely on the L1 or represent a new network of traces associated with the L2. We tried to answer this question by looking at the processing differences for the prepositions separately for the three participant groups.

In our further analyses, we were particularly interested in the differences between the three prepositions. For L1 speakers of German, we found robust differences between *auf* and *über*. One possibility is that these differences reflect differences in the way these prepositions carve up the upper space. While *über* refers to objects that do not touch the relatum, *auf* refers to objects that are in direct contact with the relatum. Therefore, we can expect to find stronger compatibility effects for *über* than for *auf*. Indeed, we did find a compatibility effect for *über* but not for *auf* in German L1 speakers. We initially expected L2 speakers with a Russian or English background to show the exact same effect pattern as the German L1 speakers, due to the preposition similarities between the languages. But as seen in Experiment 2, the L2 speakers showed compatibility effects for *auf* as well as for *über*. It seems that they have experiential traces reactivated for all stimuli, showing that they see *auf* and *über* as spatially related. One possible reason for this difference to German native speakers might be that German L1 speakers first learn *auf* as a spatial term, but in German the word *auf* is also frequently used in various contexts in which the spatial meaning is no longer obvious (e.g., *aufhören 'to stop/ terminate', aufmachen 'to open sth.', sich auf etwas freuen 'to look forward to sth.'*). It therefore appears possible that the spatial meaning fades away over the life course as people gain more non-spatial experiences with this word, which might explain why there was no compatibility effect for *auf* in the German native speakers. However, this hypothesis needs further testing before definite conclusions can be drawn. One possibility would be to compare adult language processing with that of children who are just starting to learn these words and therefore have different linguistic experiences than adults. Another possibility would be a control study where the spatial meaning would be clearly triggered, for example by using prepositional phrases.

In Experiment 3, we tested L2 speakers of German with Korean or Turkish as L1. We expected them to show a different processing pattern than German L1 speakers, as their languages do not distinguish between contact and noncontact along the vertical axis and are therefore not directly connectable to German words. Indeed, we found clear processing differences. L2 speakers with Turkish or Korean as L1 showed compatibility effects for *auf* but not for *über*, quite the opposite of the German L1 speakers. One possible reason for the reversal of this effect lies in the frequency of this word. *Auf* is much more frequent than *über* in learners' input (Bryant, 2012), as well as in total use (e.g., Quasthoff, Fiedler, & Hallsteinsdóttir, 2011). In terms of different word formation products with their respective morphemes, many more words with the prefix *auf-* (5591 word form entries) exist than with the prefix *über-* (4065 word form entries; see Quasthoff et al., 2011). These frequency differences might account for the reversal of the effect, as L2 speakers search for an equivalent to their spatial terms for the upper dimension and the first spatial term they learn is *auf*, which is then taken as the sought-for equivalent. *Über*, which is learned later, is then harder to connect, in particular because the category of spatial relations to be split in this case.

To tentatively test this hypothesis, we compared mid/low and highly proficient participants in Experiment 3, and indeed found a difference in the expected direction: Mid-to-low proficient participants showed compatibility effects for *auf* but not for *über*, while highly proficient participants showed compatibility effects for *über* as well as *auf*. However, as the sample sizes were rather small after the subcategorization, we cannot draw stable inferences from these analyses. More research is needed to investigate these points in detail.

Taken together, we found evidence for embodiment in L1 and in L2 processing. Experiential traces were reactivated in all tested groups. Furthermore, our results indicate that the observed processing differences between the groups depended on participants' native language, as hypothesized by Slobin (1996). In addition, we found evidence that language proficiency is relevant, as was also suggested by De Grauwe et al. (2014) and Vukovic (2013).

2.2.2 Detailed description of methods and results of Study 2

In the following section, details about the applied methods and obtained results for the three experiments within Study 2 are described.

Method of Experiment 1: L1 German

In this experiment, German native speakers were presented with four German words, three of which were prepositions referring to the upper or lower dimension. Their task was an adaptation of the so-called Stroop paradigm: correctly responding to the words required either an upward or downward movement depending on the font colour of the stimuli.

Participants. 49 German native speakers (11 male; M_{age} = 22.9 years, SD_{age} = 4.4 years) participated for course credit or financial reimbursement after signing a form of consent. All participants were students at the University of Tübingen. The participants had normal or corrected-to-normal vision.

Materials and apparatus. Three German words served as stimuli, namely *über* (*above*), *auf* (*on*), and *unter* (*below*). Within this, *auf* and *über* served as referents for the upper dimension; and *unter* served as referent for the lower dimension. The preposition *unter* often comes to mind as a counterpart for the preposition *über* (*über und unter*). For counterbalancing purposes, we wanted to have counterparts for both *über* and *auf*. We therefore included the word *ab* (*off/down*) as a filler word in our experimental task because this word often comes to mind as a counterpart for the preposition *auf* (*auf und ab*). However, it is important to note that the particle *ab* is not a spatial preposition, but rather combines productively with verbs without expressing the meaning of down. Consider, for example: *abwischen* (*to wipe*), *ablehnen* (*to reject*), *abmachen* (*to arrange*). Its only spatial usage is as part of the directional adverb *abwärts* (*downwards*) which probably also explains why it comes to mind as a counterpart of auf. We thus do not include the word *ab* in our analyses.

The four words were presented in four different font colours: blue (rgb 0, 0, 255), orange (rgb 255, 128, 0), lilac (rgb 150, 0, 255), or brown (rgb 140, 80, 20). Each word appeared equally often in each colour. By using four colours rather than two we made the task more complex. We wanted to make sure that participants had more time to process the word before responding. Otherwise, colour processing might already be finished before the meaning of the word had been accessed (cf. conflict monitoring theory, Botvinick, Braver, Barch, Carter, & Cohen, 2001). This is a common procedure in Stroop-like tasks (e.g. Lachmair et al, 2011; Dudschig, de la Vega, & Kaup, 2015). Responses were recorded using a PS/2 computer-keyboard adapted with a locally constructed overlay (Figure 5).

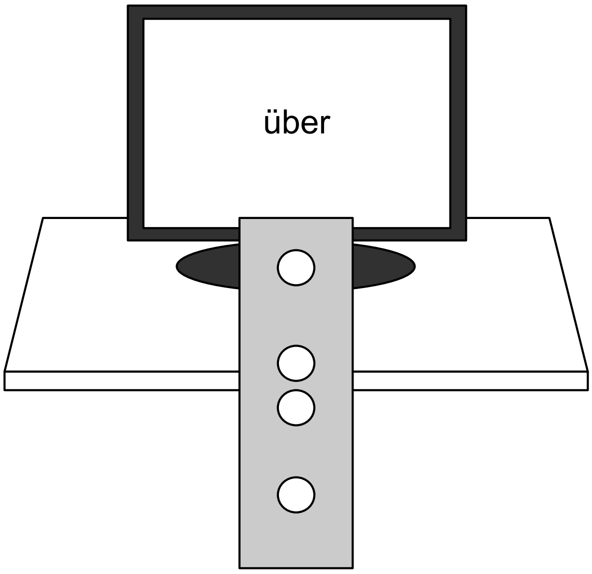


Figure 5. Experimental setup. The keyboard is implemented under a vertical plane in front of the participants. A response is made by releasing one of the middle buttons, pressing a button above or below, and returning back to the released middle button, while the other hand rests on the respective middle button.

Procedure and design. Each trial started with a fixation cross, displayed in the centre position of the screen for 1000 ms. Afterwards, the stimulus was presented at the same position as the fixation cross until the participant responded by releasing one of the middle buttons. After button release, a blank screen was shown until the corresponding upper or lower button was pressed. Between trials, a white screen was shown for 1000 ms.

Every word was presented 80 times, resulting in a total amount of 320 trials, which were subdivided into 4 experimental blocks. The experiment started with a practice block in which 16 stimuli were presented two times each in different colours. These stimuli were different from the experimental stimuli. In contrast to the experimental blocks, the participants received feedback about response accuracy during the practice block.

As can be seen in Figure 5, participants used a response box with four buttons for the task. At the beginning of each trial, they were asked to push down the two middle buttons and keep their hands there until responding. Half of the participants used the right hand for the upper middle button and the left hand for the lower middle button. For the other half of participants, this mapping was reversed. Participants were instructed to respond to the font colour of the stimuli as quickly and accurately as possible by means of an upward or downward arm movement. That is, participants had to release the respective middle button and press the upper or lower button, depending on the colour of the presented word, before returning to the middle button. The upper and lower buttons were each associated with two of the four possible colours. This mapping of colours to response direction was balanced across participants. All possible colour pairs occurred equally often and were randomly assigned to the two buttons. Unlike in the traditional Stroop task, the response direction indicated by the font colours, not the font colour itself, defined the compatibility conditions (compatible condition: e.g., upward response to *über*, incompatible condition: e.g., downward response to *über*). This is a common practise in many Stroop-like experiments nowadays (for an overview see MacLeod, 1991).

The design was a 3 (stimulus: *auf*, *über*, *unter*) x 2 (response direction: upward vs. downward) within-subjects design. The dependent variable was the release time of the middle button.

Results and Discussion

Two participants were excluded from data analysis due to an error rate above 15%. The mean error rate was 3.9%. Release responses faster than 200 ms and slower than 3000 ms and error trials were excluded from further analyses. Responses deviating by more than 3 SDs from the mean for each participant and condition (stimulus x response) were excluded. Outlier elimination reduced the data by 1.6%. Mean response times are displayed in Figure 6. We do not display the mean response times for the filler word *ab*, as it was not included in our analysis. However, as expected the response times for downwards responses (560 ms) and upwards responses (563 ms) on *ab* did not differ significantly from each other $t(46) = 0.36$, $p = .722$.

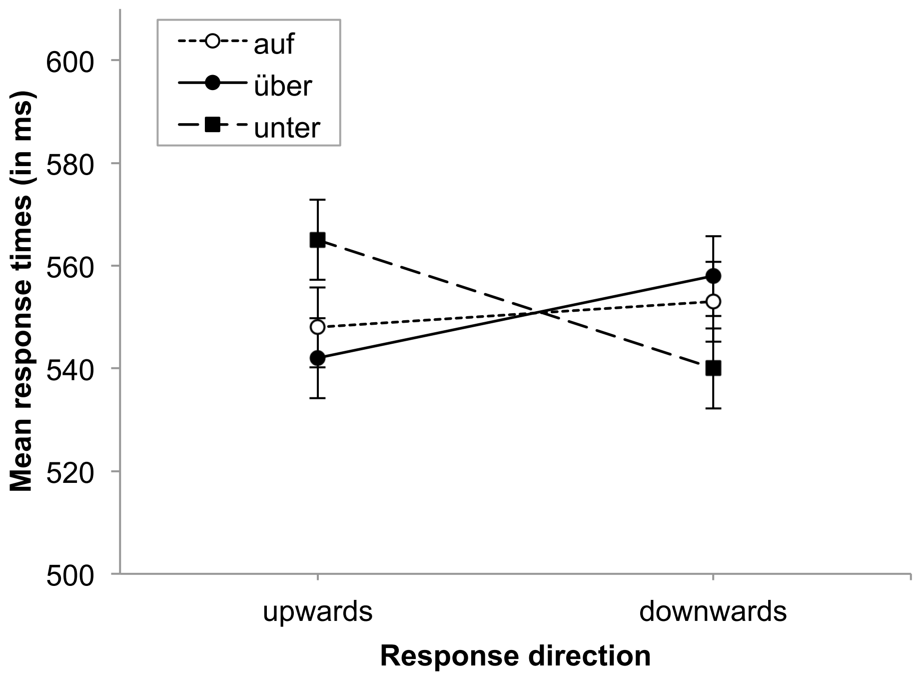


Figure 6. Mean response times and standard errors among German L1 speakers for correct responses as a function of response direction and stimulus. Error bars represent 95% confidence intervals (as per Masson & Loftus, 2003).

The analyses revealed a significant compatibility effect, represented by a significant interaction between stimulus and response direction, $F(2, 92) = 14.38$, $p < .001$, $\eta_p^2 = .238$. The main effects for stimulus, $F(2, 92) = 0.35$, $p = .706$, $\eta_p^2 = .008$, as well as response direction, $F(1, 46) = 0.05$, $p = .830$, $\eta_p^2 = .001$, were not significant. To gain more information concerning the significant interaction, we looked into the differences between compatible and incompatible response times for the different stimuli separately. We found a significant difference for the words *über*, with faster response times for upwards responses compared to downward ones, $t(46) = -2.07$, $p = .045$, as well as a significant difference for *unter* with faster response times for downwards responses than upwards ones, $t(46) = 3.04$, $p = .004$, but no significant difference for the word *auf*, $t(46) = -0.69$, $p = .496$.

These results nicely show that the Stroop paradigm is suitable for the investigation of spatial prepositions, as the observed compatibility effect is similar to the one obtained for nouns referring to entities with a typical location in vertical space (*airplane* vs. *worm*, see above). Additionally, the results imply that experiential traces even become activated when people read small units like prepositions, which can be considered further evidence in support of embodied theories of comprehension. This finding is in line with previous research using subliminal presentations of spatial prepositions (Ansorge, Kiefer, Khalid, Grassl, & Koenig, 2010) and with research on the Simon effect in response to spatial words (Khalid & Ansorge, 2013). One might argue that these results could also be explained by assuming that participants internally verbalized the response direction when planning the response in the present paradigm (i.e., verbalizing "upwards" when detecting a particular font colour) and that this caused the compatibility effect with the meaning of the presented preposition. However, a recent study conducted in our lab indicated that language-space compatibility effects are observed even in a modified Stroop-paradigm where there is no stable mapping of colours to response directions, and inner speech can therefore be ruled out as the main contributing factor (Dudschig & Kaup, 2016).[6] We therefore can be quite sure that the compatibility effect observed in this experiment (interaction of stimulus and response direction) indeed constitutes an embodiment effect and reflects the automatic activation of experiential traces during word processing.

We will now turn from discussing the overall compatibility effect to differences in the processing of *auf* and *über*. As predicted, German native speakers processed *auf* and *über*

[6] More specifically, in this experiment, participants saw words in the centre of the screen in four different font colours, as well as four coloured rectangles, located above, below, and to the left, and right of the word stimulus. The participants' task was to move towards the rectangle that matched the font colour of the word stimulus, and the location of the coloured rectangles randomly changed from trial to trial. Participants could therefore not improve the ease of their responses by memorizing a rule such as "red is upwards," making it therefore highly unlikely that inner speech is responsible for the observed compatibility effects.

differently, even though both words refer to the upper dimension. As predicted, it seems that *über* is more strongly connected with spatial experiential traces than *auf*, as for *auf* the attributed spatial dimension neither interfered with nor facilitated the response. In light of these results, it is now interesting to look at responses of L2 speakers of German with an L1 that has the same split in the upper dimension and who can therefore be expected to show the same pattern of results as the German native speakers.

Method of Experiment 2: L1 Russian or English

In Experiment 2, we focused on participants who learned German as a second language and whose native language is similar to German with respect to the upper dimension, namely Russian and English.

Participants. Forty-eight speakers with German as L2, all students or employees of the University of Tübingen, participated in this study. Twenty-four of these had English as their L1 and 24 had Russian as their L1. All participants received course credit or financial reimbursement for their participation. Three participants needed to be excluded from the sample, as they were German/English bilinguals. They had learned German from their parents or other family members before entering kindergarten. For an overview of the distribution of the ages of acquisition (M_{AoA} = 15.9 years, SD_{AoA} = 5.6 years) and the language proficiencies ($M_{proficiency}$ = 5.0, $SD_{proficiency}$ = 1.1.) of the remaining participants, see Table 2. Thirty-one of the remaining 45 participants were female and 14 were male (M_{age} = 25.9 years, SD_{age} = 4.5 years). All participants had normal or corrected-to-normal vision. They were asked to sign a form of consent before participation.

Table 2: *Distribution of Language Proficiency and Age of Acquisition*

	L1 Russian or English
Age of Acquisition	
Early L2 acquisition (3-6)	0
L2 acquisition as children (7-12)	16
L2 acquisition as teenagers and adults (12<)	29
Language Proficiency - CEFR-Levels	
A2	2
B1	3
B2	7
C1	16
C2	17

Note. The stages of language acquisition were adapted from Klein (1992); Language proficiency was measured on a 6-point scale referring to the CEFR levels (Common European Framework of Reference; Verhelst, Van Avermaet, Takala, Figueras, & North, 2009)

Materials and procedure. The stimuli and experimental procedure were identical to Experiment 1. In addition, the participants in this experiment received a short questionnaire after the main study concerning their language background and proficiency. We assessed how many foreign languages they had learned, age of acquisition of German, and their subjective evaluation of their language proficiency on a 6-point-scale referring to the levels of the Common European Frame of Reference (CEFR, Verhelst et al., 2009; see Table 1 for an overview). According to the CEFR, language proficiency is divided into 6 categories: A1, A2, B1, B2, C1, and C2. The two A levels refer to the learner as a basic user who interacts with natives in a very simple way and can use basic expressions and later on phrases or describe routines. The B levels see the learner as an independent user who can understand main points of standard conversations and later on texts. At this stage, he becomes able to interact with natives in a more spontaneous and fluent way. The C levels describe a proficient user, who can use the language flexibly and effectively for social, academic, and professional purposes. In the latest stage, the user understands virtually everything that is heard or read, and is even able to use idiomatic expressions (Verhelst et al., 2009).

Results and Discussion

Data were analysed as in Experiment 1. Outlier elimination reduced the dataset by 1.8%. The mean error rate was 4.2%. Mean response times for *auf*, *über*, and *unter* are displayed in Figure 7. As expected, The response times for *ab* in this language group also did not differ significantly between upward (653 ms) and downward responses (654 ms), $t(44) = -0.088$, $p = .930$.

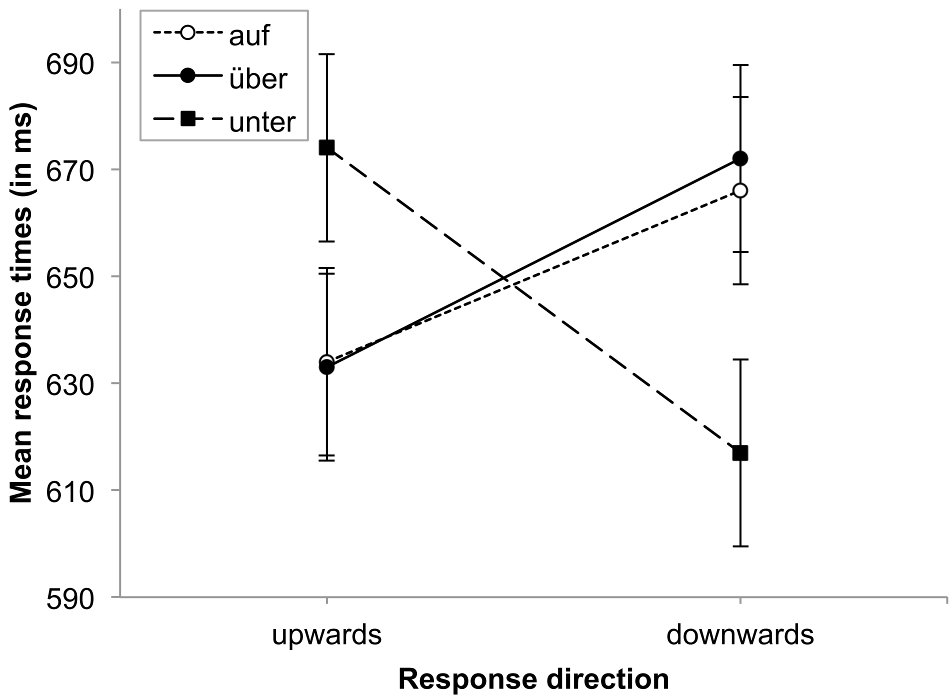


Figure 7. Mean response times and standard errors of Russian and English L1 and German L2 speakers for correct responses as a function of response direction and stimulus. Error bars represent 95% confidence intervals (as per Masson & Loftus, 2003).

The analyses again revealed a significant compatibility effect, represented by a significant interaction between stimulus and response direction, $F(2, 88) = 18.71$, $p < .001$, $\eta_p^2 = .298$. The main effects for stimulus, $F(2, 88) = 1.12$, $p = .330$, $\eta_p^2 = .025$, as well as response direction, $F(1, 44) = 0.43$, $p = .517$, $\eta_p^2 = .010$, were not significant. Separate analyses for the three stimuli revealed significant compatibility effects for all three words with shorter response times for upwards responses compared to downwards ones for *auf*, $t(44) = -4.03$, $p < .001$, and *über*, $t(44) = -2.99$, $p = .005$, and shorter response times for downwards responses compared to upwards ones for the word *unter*: $t(44) = 3.91$, $p < .001$[7]. The overall compatibility effect (interaction of stimulus and response direction) shows that compatibility effects can be observed during L2 processing, therefore supporting the view that experiential traces are reactivated during second language processing. This overall effect, however, leaves open whether the reactivated traces stem from L1 or L2 use. More information with respect to this question can be obtained when comparing the observed pattern with that observed for German native speakers. Interestingly, the pattern of results differed from that of the German native speakers. Whereas German native speakers showed no effect for *auf*, the participants in

[7] The English *under* and the German *unter* are cognates, while the Russian *под* and German *unter* are non-cognates. Therefore, it could have been possible that Russian L1 speakers and English L1 speakers differ in their responses to this word. We checked and found no significant differences between the English and Russian L1 speakers with $F < 1$ in the general analysis as well as in a separate analysis for *unter* only.

this experiment did show an effect for *auf*, suggesting that they processed *auf* spatially. More research is needed to investigate in what way the native language of the speaker accounts for this kind of processing difference. Possibly, the observed effect for *auf* in these speakers results from the fact that the equivalent of *auf* in their first language is used spatially to a stronger degree than *auf* in German. This interpretation is in line with the coactivation account (Blumenfeld & Marian, 2013; Kroll, Bobb, & Hoshino, 2014), according to which bilinguals always activate both languages (see above).

Method Experiment 3: L1 Korean or Turkish

The purpose of Experiment 3 was to compare the results of Experiments 1 and 2 to participants who learned German as L2 and whose native language shows no category split of the upper subspace into +/- contact, as is the case in Korean and Turkish.

Participants. Fifty-two students or employees of the University of Tübingen with German as L2 participated in this study. Twenty-four of these had Korean as their L1 and 28 had Turkish as their L1. All participants received course credit or financial reimbursement for their participation. Nine of the Turkish native speakers had to be excluded, as they were German/Turkish bilinguals. They had learned German from their parents or other family members before they entered kindergarten. The ages of acquisition of German (M_{AoA} = 14.5 years, SD_{AoA} = 6.8 years) and the language proficiency ($M_{proficiency}$ = 4.2, $SD_{proficiency}$ = 1.4) for the remaining 43 participants can be seen in Table 3. The language proficiency of this group of participants did not differ significantly from the participants in Experiment 2: $F(1, 85) = 2.96$, $p = .089$. Thirty-five of the participants were female and 8 male (M_{age} = 24.7 years, SD_{age} = 5.8 years). All participants had normal or corrected-to-normal vision. They were asked to sign a form of consent before participation.

Table 3: *Distribution of Language Proficiency and Age of Acquisition*

	L1 Korean or Turkish (N)
Age of Acquisition (Age Range)	
Early L2 acquisition (3-6)	9
L2 acquisition as children (7-12)	3
L2 acquisition as teenagers and adults (12<)	30
Language Proficiency - CEFR-Levels	
A2	6
B1	6
B2	15
C1	3
C2	12

Materials and procedure. The stimuli as well as the experimental procedure were identical to Experiment 2.

Results and Discussion Experiment 3

Data were analysed as in Experiments 1 and 2. One participant with L1 Korean was excluded from the data analysis due to an error rate above 15%. The mean error rate was 4.3%. Outlier elimination reduced the data set by 2.1%. Mean response times for *auf*, *über*, and *unter* are displayed in Figure 8. The response times for *ab* for upward responses (620 ms) and downward responses (648 ms) did differ significantly in this language group, $t(39) = -2.30$, $p = .027$. This pattern was different in comparison to the two language groups reported above. However, as of yet we have no explanation for this response time difference.

As in the previous two experiments, the analyses revealed a significant compatibility effect, represented by a significant interaction between stimulus and response direction, $F(2, 82) = 18.29$, $p < .001$, $\eta_p^2 = .309$. The main effects for stimulus, $F(2, 82) = 2.9$, $p = .061$, $\eta_p^2 = .066$), and response direction, $F(1, 41) = 0.06$, $p = .804$, $\eta_p^2 = .002$, were not significant. Separate analyses for the three stimuli showed a significant compatibility effect for the word *auf*, with shorter response times for upwards responses compared to downwards ones, $t(41) = 2.82$, $p = .007$, and a significant compatibility effect for the word *unter*, with shorter response times for downwards responses than for upwards ones, $t(41) = 3.60$, $p < .001$. No significant compatibility effect was obtained for *über*, $t(41) = 1.60$, $p = .117$.

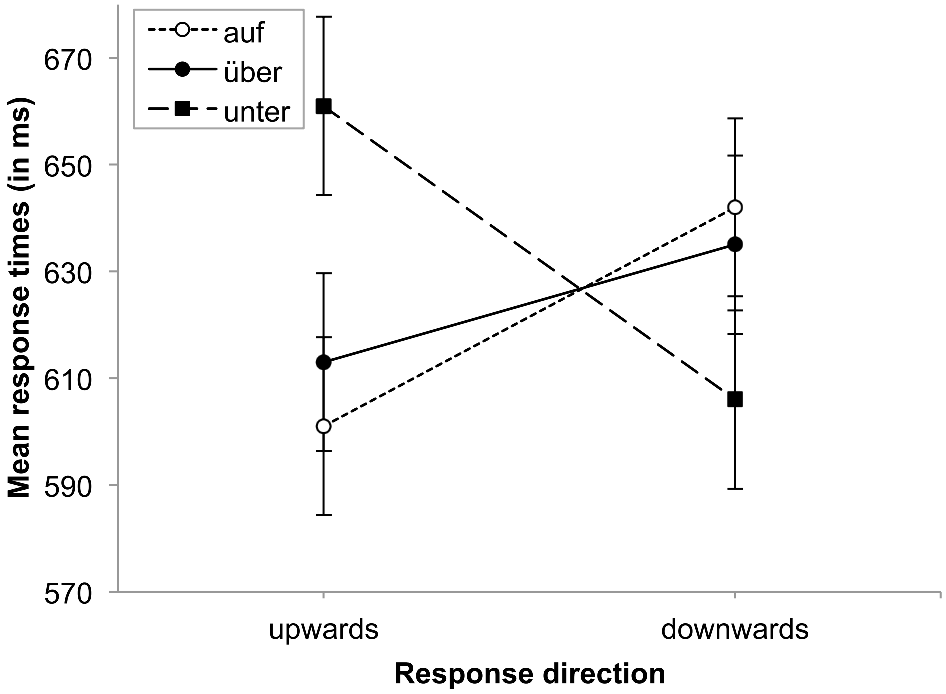


Figure 8. Mean response times and standard errors of Korean and Turkish L1 and German L2 speakers for correct responses as a function of response direction and stimulus. Error bars represent 95% confidence intervals (as per Masson & Loftus, 2003).

The results of this experiment show that Korean and Turkish native speakers process German prepositions in a different way than German native speakers. As expected, these speakers in contrast to German native speakers do not show a larger effect for *über* compared to *auf*, which presumably reflects the fact that their L1 (Korean/Turkish) does not distinguish between the two corresponding spatial relations, and they therefore do not replicate this distinction in the German input. In fact, these speakers not only do not show a larger effect for *über* compared to *auf* (as the German natives do), but quite the contrary is the case: they show a significant compatibility effect for *auf* but not for *über*. We consider it likely that this difference is related to the fact that *auf* is much more frequent compared to *über* (Ruoff, 1990) and also acquired much earlier (Bryant, 2012; Grimm, 1975). The Korean and Turkish native speakers probably simply transferred all their experiential traces related to the upper vertical space to the more frequent term *auf* and more or less ignored the less frequent *über*. If this is actually the case, one might expect to find differences between participants with higher and lower language proficiency in German. It could be that *über* leads to compatibility effects only for highly proficient L2 speakers of German because they either transfer the experiential traces of their L1 also to *über* or with enough experiences they build new traces for their L2. To investigate this question, we included language proficiency as an additional factor in the analysis of this group. We categorized the participants into two groups according to their language proficiency. Speakers with a CERF level of A2, B1, or B2 were considered as mid-to-low proficient, speakers with a CERF level of C1 or C2 were considered highly proficient. The analysis revealed a significant interaction between stimulus, response direction, and language proficiency, $F(2, 78) = 3.35$, $p = .040$, $\eta_p^2 = .077$, indicating that a difference does indeed exist between the high and mid-to-low proficiency groups. In order to gain more information with respect to our hypotheses concerning the role of language proficiency, we analysed the compatibility effects for the three stimuli and the two proficiency groups separately. For the mid-to-low proficiency group, *auf* showed a significant compatibility effect, $t(26) = -2.26$, $p = .016$ (one-tailed), while for *über* no significant compatibility effect was obtained, $t(26) = 1.06$, $p = .149$ (one-tailed). In the high proficiency group, both *auf* and *über* showed significant compatibility effects, *auf*: $t(14) = -1.77$, $p = .049$ (one-tailed); *über*: $t(14) = -1.91$, $p = .038$ (one-tailed). These results fit with the assumption that *über* is acquired later than *auf* also in L2 acquisition.

Comparing the Results of Experiment 1 – 3

In order to substantiate the described differences between the speaker groups, we conducted another analysis with L1 group as an additional between-subjects factor: 3 (L1 group: German Natives, L1 English or Russian, L1 Korean or Turkish) x 3 (stimulus: *auf*, *über*, *unter*) x 2 (response direction: upward vs. downward).

Our analysis also showed a significant main effect for L1 group, $F(2, 131) = 6.56$, $p = .002$, $\eta_p^2 = .091$, because the German L1 speakers responded faster overall than the other L1 groups, $M_{GermanL1} = 551$ ms, $SD_{GermanL1} = 105$ ms; $M_{EnglishRussianL1} = 626$ ms; $SD_{EnglishRussianL1} = 151$ ms; $M_{KoreanTurkishL1} = 650$ ms, $SD_{KoreanTurkishL1} = 165$ ms. The German L1 group differed significantly from the L1 Korean/Turkish group, $F(2, 174) = 6.81$, $p = .001$, $\eta_p^2 = .073$, as well as from the L1 Russian/English group, $F(2, 180) = 6.22$, $p = .002$, $\eta_p^2 = .065$. The two groups with German as L2 did not differ significantly from each other, $F < 1$. The interaction between L1 group and stimulus was marginally significant, $F(4, 262) = 2.23$, $p = .067$, $\eta_p^2 = .033$. The other main effects, as well as the interaction between response direction and L1 group, were not significant (Fs < 1). As expected, the interaction between stimulus and response direction was significant just as in the separate experiments, $F(2, 262) = 48.59$, $p < .001$, $\eta_p^2 = .271$. Most importantly, and also as predicted, we obtained a significant 3-way-interaction between L1 group, stimulus, and response direction, $F(4, 262) = 3.84$, $p = .005$, $\eta_p^2 = .055$, supporting our interpretation that German spatial prepositions are processed differently by German native speakers and different groups of L2 speakers of German.

2.3 Study 3 - Language motor interactions in bilingual school-children

Reference

Ahlberg, D.K., Bischoff, H., Strozyk, J.V., Bryant, D., & Kaup, B. (2016). *How do German bilingual schoolchildren process German prepositions? – A study on language-motor interactions.* Manuscript submitted for publication.

2.3.1 Summary of Study 3

In our previous study, we tested adults who mainly learned their second language after the age of twelve. Therefore, the semantic categorizations of the L1, as well as the connected experiential traces were probably already very strongly consolidated. As it is well known that the age at which a language is being acquired plays an important role in the acquisition process (Hyltenstam & Abrahamson, 2003; Johnson & Newport, 1989; Meisel, 2009), in the current study we aimed at investigating compatibility effects in bilingual participants who learned German as well as at least one other language (OL) before the age of six. We were interested in whether these bilinguals process German spatial prepositions in a similar way as German monolinguals do, or whether their processing is influenced by the nature of the OL. To investigate this, we presented participants with a Stroop-like task similar to the one we used in our adult study (Study 2), concentrating on the prepositions *auf*, *über* and *unter*. We presented these stimuli in four different font colours. A correct response was made by either an upward or downward movement depending on the font colour of the stimuli. Font colour was thus the imperative stimulus in the present task. We compared three groups of children: The first group of children had only acquired German until the age of six. The second group of children had acquired German as well as at least one other language until the age of six, whereby the other language/languages were similar to German with respect to the division of the upper subspace. The third group of children had acquired German as well as at least one other language until the age of six, whereby the other language/languages were dissimilar to German with respect to the division of the upper subspace.

Just as in the experiment with adult participants, we expected to find compatibility effects (i.e., faster responses for compatible compared to incompatible trials) in all groups. Children in secondary schools should have already developed a network of experiential traces for spatial categories, as well as sufficient reading fluency (Günther, 1986), to show similar compatibility effects as adults. With regard to the processing of the individual prepositions, different scenarios are imaginable. First, semantic categorizations in the OL might be predominant despite the fact that German was acquired at an early age and participants have had several years of language contact with German. If so, it can be expected that the spatial system of the other language has an impact on the processing of spatial terms in German. In this case we expect different compatibility effects for different groups of speakers depending on the nature of their OL, just as in our study with adult participants. Second, early age of acquisition and several years of contact with German may have allowed the children to develop an independent spatial system of German even if it deviates from the spatial system in their OL. If so, we expect to find comparable compatibility effects for all groups of children, independent of the nature of their OL.

A further point we wanted to investigate is the role of language proficiency. Some authors hypothesize that language proficiency plays a major role in the development of experiential traces (De Grauwe et al., 2014; Vukovic, 2013). Furthermore, in Study 2 with adult participants, we found tentative evidence that experiential traces might change over time with increasing proficiency. For instance, while for highly proficient participants a compatibility effect was found for *auf* as well as *über*, in the mid-to-low proficient participants the compatibility effect was only found for *auf*, not for *über*. Thus, the pattern of results for highly proficient participants resembled the pattern of results of native speakers more closely than that of the mid to low proficient participants. This finding is in line with the findings of Bryant (2012), according to which *auf* is learned earlier than *über*. However, the sample size in our previous study (Study 2) was not large enough to draw stable inferences after the sub-categorization into two proficiency groups. Therefore, we added an objective measure to assess language proficiency in the present study to further investigate this question.

In the present study, the results confirmed the hypothesis that experiential traces are being reactivated during word processing in all groups. All groups showed faster responses when the meaning of the preposition was compatible with the direction of the motor response (e.g., upward movement for *auf*) compared to incompatible trials (e.g., downward movement for *auf*). This finding is in line with the results obtained for adult L2 speakers (Study 2). It seems that all participants (i.e., both German monolinguals and German bilingual children) reactivated experiential traces connected to the spatial categories in a similar way as the adults. This is of special interest, as it shows that experiential traces are already established and connected to words at the age of eleven to fifteen years, and can be automatically accessed in a task that does not require active reading.

However, contrary to the adult speakers, the children did not differ across the language groups. Our results showed that bilingual children who learned German relatively early in life processed German prepositions in the same way as German monolingual children, even when their OL uses a different spatial categorization than German. In addition, when we looked at the different proficiency levels directly, we found some subtle differences between highly and lowly proficient speakers. However, as the proportion of children showing a childlike L2 acquisition of German was less than 10%, we were not able to compare them directly to the early bilinguals (i.e., the simultaneous and early-successive bilinguals). I will further discuss this point more concretely in the general discussion.

Another point worth mentioning is the fact that the effects for German native speakers were quite different for the children compared to the adults of Study 2. Contrary to the adults, who showed a compatibility effect for the word *über*, but not for the word *auf*, the children showed a compatibility effect for *auf*, but not for *über*. One possible explanation for this finding is that the word *auf* is much more frequent in the early learner input. Therefore, children might access the meaning of the word *auf* earlier during the comprehension process than the meaning of the word *über* (Bryant, 2012). However, it is surprising that we still found this effect for eleven to fifteen years olds. Although these children have more than ten years of German language experience, they still show this effect. I will further discuss this point including the possibility of a restructuring process in the general discussion.

In sum, we found compatibility effects for spatial prepositions in a Stroop-like task among monolingual and bilingual children. Thus, we could confirm that the experiential traces account can be applied to language processing in children as well as to language processing in bilinguals. In addition, this study provides a good starting point to further investigate processing differences between early (before the age of six) and late (after the age of twelve) language learners as well as between highly and lowly proficient speakers of German with respect to experiential traces.

2.3.2 Detailed description of method and results of Study 3

In the next section I will describe the applied method and the obtained results of Study 3.

Method

Participants. Three-hundred-eighty-three schoolchildren at different secondary schools in Southern Germany took part in our experiment. They received financial reimbursement on a class basis for their class treasury. The experimental testing was in agreement with the guidelines for good scientific practice at the LEAD Graduate School at the University of

Tübingen (Germany). This was checked and approved by the ethics committee at the Faculty of Economics and Social Sciences, University of Tübingen. Prior to experiment participation the parents of our participants gave their written informed consent. Throughout the data acquisition the data were connected only by a participant code and at no point could the recorded data be associated with a participant's name.

We grouped our participants into three groups: children who had only learned German as a native language and no additional language until the age of 6 (in the following: "German monolinguals"); children who learned German and at least one other language until the age of six, whereby the other language or the other languages split up the upper subspace with two different expressions, just like German (e.g., Russian, English, Italian; in the following: "German bilinguals: similar OL"); children whose first language does not further distinguish the upper subspace (e.g., Turkish, Urdu, Swahili; in the following: "German bilinguals: dissimilar OL"). For an exact overview of the language group assignment, see the supplementary material. We needed to exclude 19 participants from our sample because we were not able to categorize them into one of the above groups. Either we were not fully able to tell whether their OL is similar to or dissimilar to German due a lack of information from the participants about which dialect from a particular country they spoke (e.g., Eritrea, Nigeria), or the categorizations of their other languages conflicted with one another (e.g., English and Turkish; Turkish and Kurdish). In addition, we excluded 53 participants who committed errors on more than 20% of the trials, and 10 participants who responded in less than 100 ms on more than 20% of the trials. Although clearly instructed to use only their dominant hand to respond, some children could not be prevented from using both hands in the experiment. This led to the attainment of response times lower than 100 ms. To be sure to include only children who followed the instructions, we used this as an exclusion criterion.

The remaining 320 participants (M_{age} = 13.0 years, SD_{age} = 1.5 years, 166 male, 289 right-handed) were distributed over our three groups as follows: 130 German monolinguals (M_{age} = 13.0 years, SD_{age} = 1.5 years), 138 German bilinguals with a similar OL (M_{age} = 13.0 years, SD_{age} = 1.5 years), and 52 German bilinguals with a dissimilar OL (M_{age} = 13.1 years, SD_{age} = 1.7 years). For a more detailed overview on the distribution across class levels and school types see Table 4. All participants had normal or corrected-to-normal vision.

Table 4: *Descriptive information about class level, school type, and bilingualism of our participants*

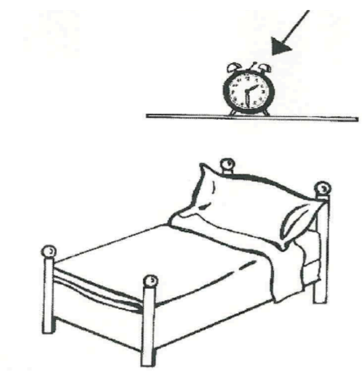
Descriptives	Language Group		
	German Mono-linguals	German Bilinguals: Similar-OL	German Bilinguals: Dissimilar-OL
Class Level (in %)			
5	10.8	18.1	19.2
6	25.4	23.9	19.2
7	16.9	17.4	11.5
8	26.2	22.5	28.8
9	20.8	18.1	21.2
School Type (in %)			
Werkrealschule	56.2	76.8	73.1
Gemeinschaftsschule	29.2	21.0	21.2
Realschule	14.6	2.2	5.8
Grade of Bilingualism (in %)			
Simultaneous Bilingualism		63.0	75.0
Early-Successive Bilingualism		11.6	11.5
Childlike L2 Acquisition		12.3	7.7
Not Specifiable		13.0	5.8

Note. The categorization of bilingualism was made according to Rothweiler and Kroffke (2006) depending on the age of acquisition of German: 0-3 years – simultaneous bilingualism; 3-5 years – early-successive bilingualism; > 5 years – childlike L2 acquisition. A few children could not be classified, as their questionnaires were incomplete. In Germany, different types of secondary schools exist. Werkrealschule and Realschule offer secondary education for years 1-10, with a stronger focus on practical skills in the Werkrealschule. Additionally, the Gymnasium qualifies for University education. Gemeinschaftsschule is a school that serves as a combination of these three school types.

Material

Language proficiency and prepositional knowledge tests. To gain objective information about the children's proficiency in German, we conducted language proficiency tests in form of a "C-Test" (Grotjahn, 1992; Grotjahn, 2014). This paper-pencil test consists of four short texts with 20 gaps (with a length of about half a word) that each needed to be filled in. The scoring gives two measures: The word-recognition score and the accuracy score. The word-recognition score represents the number of correctly recognized words, regardless of their spelling accuracy. The accuracy score additionally takes spelling into account. Both scores can reach a maximum of 80, one for each gap, whith the accuracy score typically lower than the word-recognition score. We used different tests for each class level to prevent floor or ceiling effects. The test for fifth graders was taken from Baur, Chlosta, and Goggin (2011). The remaining tests for 6^{th} to 9^{th} graders were provided by the same authors (Baur & Goggin, 2005a, 2005b; Goggin, 2011; Goggin, 2014).

To get information about whether the children have the meaning of the German prepositions used in the experiment available, we used a paper-pencil adaptation of the Topological-Relations-Picture Series (TRPS; Bowerman & Pederson, 1992) on the basis of the adaptation by Bryant (2012), who used a shortened version of this test and added different pictures to adapt it to typical German spatial configurations (see Figure 9 for an example item). We further shortened this set to 15 test pictures and one example item that was discussed with the children first. The gaps in the sentences needed to be filled with the prepositions *auf*, *über*, *unter*, *an*, and *in*, each three times. This resulted in a total maximum score of 15 and a maximum score of 9 for the prepositions *auf*, *über*, and *unter* only. For the results of both tests, see Table 5.



Der Wecker steht __________ dem Bett.

Figure 9. Example item from the adaptation of the TRPS used in this study (Bryant, 2012). The child is expected to fill in the preposition *über* (*Der Wecker steht über dem Bett / The alarm clock is standing above the bed).*

Questionnaires. We designed two questionnaires, one for the children and one for their parents. In these questionnaires we assessed the language background of the children, the origin of their parents and grandparents (country of origin and native language of parents and grandparents, as well as which languages the child speaks and how proficient he or she is in these languages), as well as the socio-economic status of the families in form of the HISEI (Ganzeboom, de Graaf, Treiman, & de Leeuw, 1992). Both questionnaires contained similar questions, but we adapted the wording and illustration to match the different target groups. By using questionnaires for children and parents we wanted to increase the possibility to get the relevant information of at least one of the groups. For the categorization and the analyses we used the data given by the parents if available, otherwise we used the data given by the children. For the assessment of the children's language contact, we included two questions. First, the children needed to indicate which languages they spoke with which persons. They were presented with a list of five persons or groups of persons (father, mother, siblings, other relatives, and friends). For each entry the children indicated which language they spoke to the respective person(s). Multiple answers were possible. We then counted how many times the children mentioned German for our German-contact measure and the amount of times the native language was mentioned for our native-language-contact variable. In the second question, we asked which languages they use in their free-time activities. Here the children were presented with six categories (watching TV, reading newspapers, reading books, reading on the Internet, listening to the radio, listening to music) and again they indicated which language or languages they used for these activities. We counted the amount of native language indications only, as due to the living environment of the children, German was used in nearly all free-time activities. A summary of the results of these questionnaires can be found in Table 5.

Table 5: *Average scores on language tests, self-reported language contact/use, and socio-economic status of the language groups*

	Language Group					
Tests	German Mono-linguals	N	Bilinguals with similar OL	N	Bilinguals with dis-similar L1	N
Language Proficiency						
Word recognition Score (%)	89.2	130	80.0	136	76.7	52
	(9.7)		(17.0)		(12.7)	
Accuracy Score (%)	79.5	130	67.1	136	62.8	52
	(13.0)		(19.4)		(16.4)	
Prepositional Use						
auf/über/unter – Score	8.4	129	7.8	138	7.5	52
(max. 9)	(1.0)		(1.7)		(1.7)	
Total Score	14.0	129	12.2	138	11.5	52
(max. 15)	(1.4)		(2.6)		(2.7)	
Language Contact						
German Contact			3.2	126	2.9	51
(1-5)			(1.3)		(1.1)	
Native Language Contact			3.4	126	3.7	51
(1-5)			(1.2)		(1.2)	
Activities in Native Language			1.5	124	2.2	49
(1-6)			(1.4)		(1.7)	
HISEI (16-90)	49.1	122	36.9	119	34.7	51
	(17.8)		(16.0)		(17.3)	

Note. Standard deviations appear in parentheses below means. For specific information regarding the acquisition of these variables, see the method section. Not all questionnaires and tests were fully completed by all participants. Therefore, we provide information about the sample size in an extra column (N).

Stimuli and Apparatus. We concentrated on three German words serving as stimuli, namely *über* (*above*), *auf* (*on*), and *unter* (*below*). Hereby, *auf* and *über* served as referents for the upper dimension which differ with respect to the feature +/- contact; *unter* served as a referent for the lower dimension. As in Study 2, we included the word *ab* (*down/off*) as a counterpart to *auf*, to balance the number of stimuli for the two dimensions. The particle *ab* is part of the directional adverb *abwärts* (*downwards*). However, as the word *ab* is neither a spatial prepositions nor is it used in explicit spatial configurations, we did not include it in our analyses but rather treated it as a filler item. Its spatial use is mostly restricted to its combination with *auf* (*auf und ab – up and down*). The four words were presented in four different font colours: blue (rgb 0, 0, 255), orange (rgb 255, 128, 0), lilac (rgb 150, 0, 255), and brown (rgb 140, 80, 20). Each word appeared equally often in each colour. Responses were recorded using a PS/2 computer keyboard adapted with a locally constructed overlay. We used LENOVO ThinkPad L530 laptops to conduct the experiment. To make it possible for the participants to view the screen despite the height of the vertically mounted keyboard, we positioned the laptops on boxes on the tables right behind the keyboards. For the exact setup, see Figure 10. The experiment was programmed with E-Prime® (Psychology Software Tools Inc., http://www.pstnet.com/E-Prime/e-prime.htm).

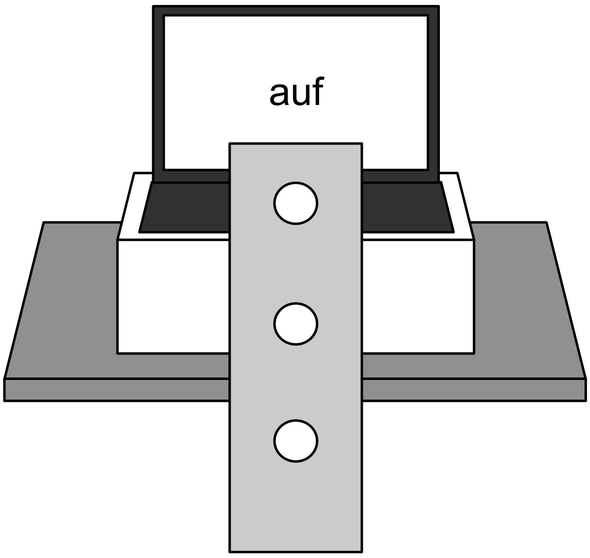


Figure 10. Experimental setup. The keyboard was implemented under a vertical plane in front of the participants. At the beginning of each trial, the participant pressed the middle key with their dominant hand. A response was made by releasing the middle button, pressing the upper or lower button, and returning back to the middle button.

Procedure and Design

In the first school, data were collected in two sessions on two separate days. However, since we lost data due to dropouts, data collections at all other schools was done in one session. We tested each child in a session of maximally 45 minutes, with 8 children in parallel. We conducted the experiment and the prepositional knowledge test right after each other in around 20 minutes of the session, while the language proficiency test took place in the remaining time. We balanced the order of those two parts; in every session, 4 children started with the experiment and 4 children did the language proficiency task. In the second half of the session, they switched tasks.

In the experiment proper, each trial started with a fixation cross, displayed in the centre position of the screen for 1000 ms. Afterwards, the stimulus was presented, also in centre position, until the participant released the middle button or for a maximum of 2000 ms. Right after the button release, a blank screen was shown until the second response, a button press of the upper or lower button, or for a maximum of 3000 ms. Between trials, a white screen was shown for 1000 ms. Please note: In the first data acquisition sessions, we showed the stimulus and the blank screen until response execution without a predetermined cut-off time. Since some children exceeded our maximum testing time of 45 minutes (determined by the length of one school lesson) in this setup, we decided to include an automatic cut-off to improve the children's motivation. For the first 93 participants, we recoded reaction times exceeding the cut-off as errors.

The participants used a response box with three buttons for their task, as can be seen in Figure 10. They were asked to only use their dominant hand (i.e., left hand for left-handers; right hand for right-handers) throughout the whole experiment. At the beginning of each trial, they were asked to push down the middle button and to keep it pressed until the stimulus appeared on the screen. When they had decided whether to press the upper or lower button depending on the font colour of the presented word, the participants were to release the middle button and press the upper or lower button instead, before returning to the middle button. The participants were instructed to respond to the font colour of the stimuli as quickly and accurately as possible.

The upper and the lower button were each associated with two of the four possible colours. This mapping of colours to response direction was balanced across participants. All possible colour pairs occurred equally often and were randomly assigned to the two buttons.

Every word was presented 40 times, resulting in a total of 160 trials, which were subdivided into 2 experimental blocks. The experiment started with a practice block consisting of 60 trials, in which we presented stimuli different from the experimental stimuli in the four colours. In both the practice block and the experimental blocks, the participants received feedback about response accuracy after each trial.

The design was a 3 (stimulus: *auf*, *über*, *unter*) x 2 (response direction: upward vs. downward) x 3 (language group: German monolinguals vs. German bilinguals with similar OL vs. German bilinguals with dissimilar OL) design with stimulus and response direction as within-subjects factors and language group as between-subjects factor. The dependent variables were the release time of the middle button as well as the errors.

Results

The mean error rate was 6.8%. For the analysis of reaction times, errors and trials with release responses or movement responses faster than 100 ms were excluded. Responses deviating by more than 3 *SD*s from the mean for each participant and condition (stimulus x response) were also excluded. The elimination of these outliers reduced the data by 2.6%. No data were excluded from the analyses of errors.

General Analysis

Reaction times. In our analysis, we obtained a significant interaction effect between stimulus and response direction, $F(2, 634) = 18.54$, $p < .001$, $\eta_p^2 = .055$, reflecting the expected compatibility effect. As can be seen in Figure 11A, responses in compatible trials (e.g., upward response to *auf*) were faster overall than responses in incompatible trials (e.g., downward response to *auf*). In addition, we found a significant main effect of response direction, $F(1, 317) = 49.42$, $p < .001$, $\eta_p^2 = .135$, with faster overall upward responses than downward responses. When looking at the individual stimuli separately, we found significant differences between compatible and incompatible responses for the prepositions *auf*, $t(319) = -8.98$, $p < .001$, and *über*, $t(319) = -5.37$, $p < .001$, but not for *unter*, $t(319) = -1.49$, $p = .136$.

The interaction between stimulus, response direction, and language group was not significant, $F(4, 634) = 1.31$, $p = .264$, $\eta_p^2 = .008$, indicating that all language groups showed similar compatibility effects. The main effects of stimulus, $F(2, 634) = 1.03$, $p = .358$, $\eta_p^2 = .003$, and language group, $F(2, 317) = 1.01$, $p = .364$, $\eta_p^2 = .006$, as well as all other interactions involving the between-subjects factor language group revealed no significant effects (stimulus x language group, $F(4, 634) = 1.52$, $p = .196$, $\eta_p^2 = .009$; response direction x language group, $F(2, 317) = 1.33$, $p = .267$, $\eta_p^2 = .008$).

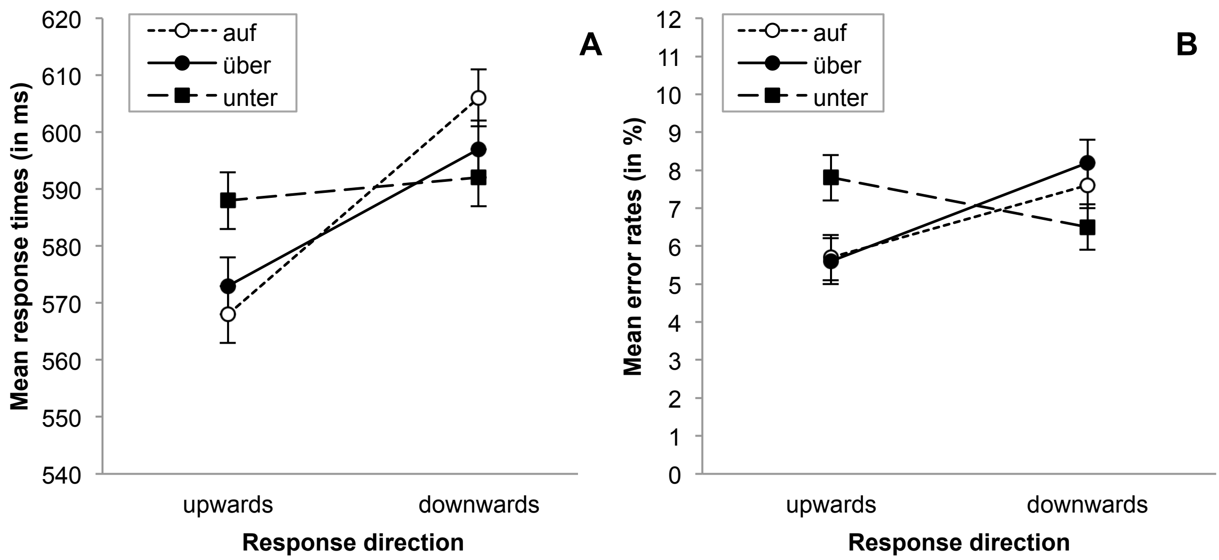


Figure 11. Mean response times for correct responses (A) and mean percentage of errors (B) as a function of response direction and stimulus for all participants. Error bars represent 95% confidence intervals (as per Masson & Loftus, 2003).

Error rates. The error analysis overall supported the results of the analysis of response times: We found a significant interaction between Stimulus and Response Direction, $F(2, 634) = 16.66$, $p < .001$, $\eta_p^2 = .050$, indicating that more errors were made in incompatible trials than in compatible trials (see Figure 11B). The main effect of response direction, $F(1, 317) = 9.73$, $p = .002$, $\eta_p^2 = .030$, was also significant: The participants made more errors on downwards trials (7.4%) than on upwards trials (6.4%). In terms of individual words, we found significant differences between compatible and incompatible responses for the preposition *auf*, $t(319) = -3.54$, $p < .001$, and *über*, $t(319) = -4.56$, $p < .001$, as well as for *unter*, $t(319) = 2.48$, $p = .014$.

Just as in the reaction times analysis, the compatibility effect did not differ depending on the language group: The interaction between stimulus, response direction, and language group was not significant, $F(4, 634) = 1.24$, $p = .291$, $\eta_p^2 = .008$. The interaction between response direction and language group, $F(2, 317) = 1.16$, $p = .314$, $\eta_p^2 = .007$, the interaction between stimulus and language group, $F(4, 634) = 1.41$, $p = .230$, $\eta_p^2 = .009$, and the main effects for language group, $F < 1$, and stimulus, $F(2, 634) = 1.00$, $p = .368$, $\eta_p^2 = .003$, were also not significant.

Taken together, the fact that we found compatibility effects in the reaction times and the error rates independent of language group implies that experiential traces got reactivated not only in the group of German monolinguals, but also in the groups of German bilinguals with similar and dissimilar OL. This is exactly what we expected. However, as we found differences between the language groups for adults (Study 2), we nevertheless conducted separate analyses for all three groups. This seemed reasonable since the main objective of our study was to investigate differences and similarities between the language groups and we wanted to make sure that we did not overlook more subtle differences.

Separate Analyses for the Different Language Groups

Reaction times. In all groups we found the same pattern of results as in the main analysis. We found a significant interaction between stimulus and response direction (German monolinguals: $F(2, 258) = 4.46$, $p = .012$, $\eta_p^2 = .033$; similar-OL: $F(2, 274) = 10.13$, $p < .001$, $\eta_p^2 = .069$; dissimilar-OL: $F(2, 102) = 6.43$, $p = .002$, $\eta_p^2 = .112$) and a significant main effect of response direction (German monolinguals: $F(1, 129) = 12.49$, $p < .001$, $\eta_p^2 = .088$; similar-OL: $F(1, 137) = 39.66$, $p < .001$, $\eta_p^2 = .225$; dissimilar-OL: $F(1, 51) = 10.12$, $p = .002$, $\eta_p^2 = .166$), while the main effect of stimulus was not significant (German monolinguals: $F < 1$; similar-OL: $F(2, 274) = 2.31$, $p = .101$, $\eta_p^2 = .017$; dissimilar-OL: $F < 1$).

Although all groups showed a similar pattern of results, Figure 12 also reveals some small differences with regard to the compatibility effects for the individual prepositions. Therefore, we compared compatible and incompatible response times for the different stimuli for each language group separately. The results can be found in Table 6. In all three groups, there was a significant compatibility effect for *auf* but not for *unter*. The only difference between the groups was that the compatibility effect for *über* was significant for both the similar and the dissimilar OL group, while it was only marginally significant for the German monolinguals.

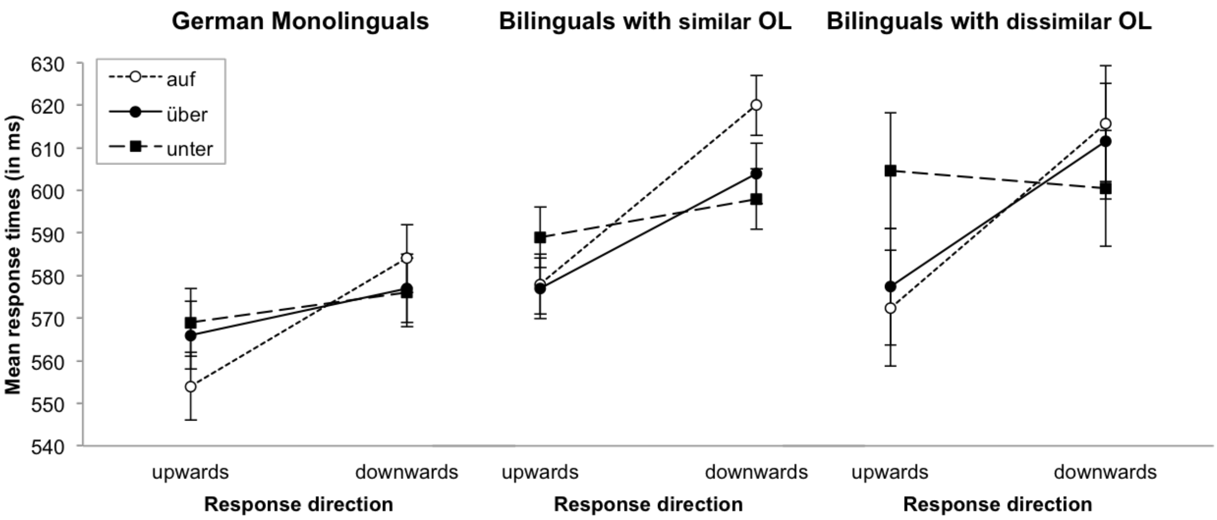


Figure 12. Mean response times for correct responses as a function of response direction and stimulus for the different language groups separately. Error bars represent 95% confidence intervals (as per Masson & Loftus, 2003).

Table 6: *Individual comparisons of the RTs in compatible and incompatible responses per language group*

		df	*t*	*p*
German monolinguals				
	auf	129	-4.43	<.001
	über	129	-1.90	.060
	unter	129	-1.00	.321
similar-OL				
	auf	137	-6.81	<.001
	über	137	-4.49	<.001
	unter	137	-1.67	.097
dissimilar-OL				
	auf	51	-4.39	<.001
	über	51	-2.88	.006
	unter	51	0.32	.748

Error rates. The analysis of the errors generally supported the findings of the reaction times analysis, as the interaction between stimulus and response direction was significant in all groups (German monolinguals: $F(2, 258) = 3.70$, $p = .026$, $\eta_p^2 = .028$; similar-O1: $F(2, 274) = 8.21$, $p < .001$, $\eta_p^2 = .057$; dissimilar-L1: $F(2, 102) = 5.58$, $p = .005$, $\eta_p^2 = .099$), whereas the main effect of stimulus was not significant (German monolinguals: $F(2, 258) = 1.93$, $p = .147$, $\eta_p^2 = .015$; similar-OL: $F < 1$; dissimilar-OL: $F < 1$). The main effect of response direction was significant for both the similar- and the dissimilar-OL groups (similar-OL: $F(1, 137) = 4.11$, $p = .044$, $\eta_p^2 = .029$; dissimilar- OL: $F(1, 51) = 4.85$, $p = .032$, $\eta_p^2 = .087$) but not for the German monolinguals($F < 1$). Mean error rates for the different language groups are displayed in Figure 13.

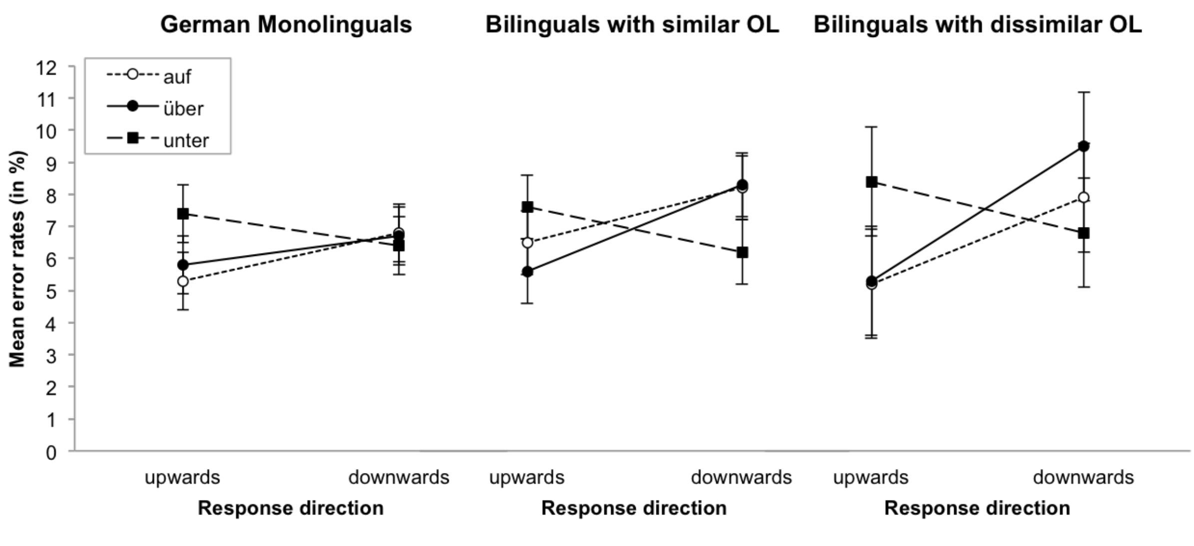


Figure 13. Mean percentage of errors as a function of response direction and stimulus for the different language groups separately. Error bars represent 95% confidence intervals (as per Masson & Loftus, 2003).

Table 7: *Individual comparisons of the error rates in compatible and incompatible trials per language group and individual word*

		df	*t*	*p*
German monolinguals				
	auf	129	-2.12	.036
	über	129	-1.17	.244
	unter	129	1.40	.166
similar OL				
	auf	137	-2.10	.038
	über	137	-3.81	<.001
	unter	137	1.64	.104
dissimilar OL				
	auf	51	-1.99	.052
	über	51	-3.12	.003
	unter	51	1.25	.218

The results of the separate analyses of the compatibility effects for the different stimuli and language groups can be found in Table 7. For the German monolingual speakers, significant differences were obtained only for the word *auf*, but not for the words *über* or *unter*. For the other two groups, the differences between compatible and incompatible responses were significant for *über*, but not for *unter*. Additionally, significant differences were obtained for the word *auf* for the similar-OL group, while the difference was only marginally significant for the dissimilar-OL group.

Analysis of Language Proficiency

As mentioned above, a second objective of the current study was to investigate the role of language proficiency in bilingual language processing of prepositions. For that reason, we pooled both bilingual groups together and conducted a median-split based on the word-recognition percentage score of the C-Test (N = 190; two C-Tests were missing, see method section). The lowly proficient bilinguals showed a mean language proficiency of 68% (*SD* = 14.7%), while the highly proficient bilinguals had a mean language proficiency of 91% (*SD* = 4.4%) on the word-recognition score. The resulting proficiency groups were included in the analysis as a between-subjects factor, resulting in a 3 (stimulus: *auf*, *über*, *unter*) x 2 (response direction: upward vs. downward) x 2 (proficiency: high vs. low) design. We did not distinguish between the similar- and dissimilar-OL groups in this analysis to increase power and because the groups of similar-OL and dissimilar-OL speakers did not differ significantly from each other in a preliminary analysis (interaction between stimulus, response direction, language group, and language proficiency, $F(2, 368) = 1.37$, $p = .254$, $\eta_p^2 = .007$). For the mean reaction times and mean error rates of the two proficiency groups, see Figures 14 and 15.

Reaction times. The main analysis revealed no significant interaction between proficiency, stimulus, and response direction, $F(2, 372) = 2.39$, $p = .093$, $\eta_p^2 = .013$, but a significant interaction between stimulus and response direction, $F(2, 372) = 17.67$, $p < .001$, $\eta_p^2 = .087$, a significant main effect of response direction, $F(1, 186) = 48.44$, $p < .001$, $\eta_p^2 = .207$, and a significant interaction between proficiency and response direction, $F(1, 186) = 4.71$, $p = .031$, $\eta_p^2 = .025$. The interaction effect for proficiency and stimulus, $F(2, 372) = 1.87$, $p = .156$, $\eta_p^2 = .010$, as well as the main effect of stimulus, $F(2, 372) = 1.87$, $p = .156$, $\eta_p^2 = .010$, and the main effect of proficiency, $F < 1$, were not significant.

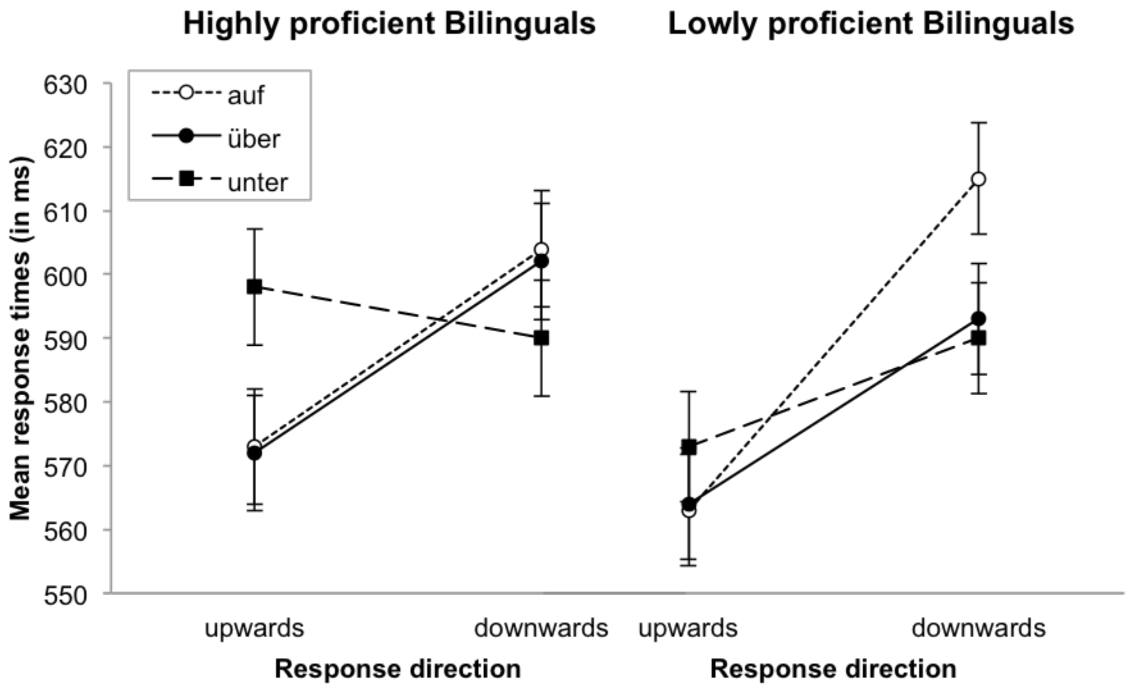


Figure 14. Mean response times for the high-proficiency and the low-proficiency group for correct responses as a function of response direction and stimulus. Error bars represent 95% confidence intervals (as per Masson & Loftus, 2003).

When we looked at the high-proficiency and the low-proficiency groups separately, we found an interaction effect for stimulus and response direction in both groups (high-proficiency group: $F(2, 182) = 11.84$, $p < .001$, $\eta_p^2 = .115$; low-proficiency group: $F(2, 190) = 8.03$, $p < .001$, $\eta_p^2 = .078$). The main effect of response direction was also significant in both groups (high-proficiency group: $F(1, 91) = 11.67$, $p < .001$, $\eta_p^2 = .114$; low-proficiency group: $F(1, 95) = 41.10$, $p < .001$, $\eta_p^2 = .302$), whereas the main effect of stimulus was not (high-proficiency group: $F(2, 182) = 1.21$, $p = .299$, $\eta_p^2 = .013$; low-proficiency group: $F(2, 190) = 2.44$, $p = .090$, $\eta_p^2 = .025$).

However, whereas in the low-proficiency group significant differences between compatible and incompatible responses were measured for all prepositions, in the high-proficiency group the differences between compatible and incompatible responses were only significant for the prepositions *auf* and *über*, but not for *unter* (see Table 8). Interestingly, as can be seen in Figure 6, the direction of the difference between compatible and incompatible responses to the word *unter* changed. For the low-proficiency group, compatible responses were slower than incompatible responses, whereas for the high-proficiency group this difference disappeared. We also conducted the same type of analysis with the variable language contact (i.e., the number of person groups with whom the children reported to communicate in German, as a group separator). We obtained very similar results as for the language-proficiency analysis, which is not surprising, as the two factors are indeed significantly correlated $r_s = .28$, $p < .001$. Thus, children who speak German with more person groups are also more proficient in German.

Table 8: *Single comparisons of the RTs in compatible and incompatible trials per proficiency group*

		df	*t*	*p*
High-Proficiency Group				
	auf	91	-4.22	<.001
	über	91	-3.97	<.001
	unter	91	1.21	.230
Low-Proficiency Group				
	auf	95	-7.19	<.001
	über	95	-3.79	<.001
	unter	95	-2.57	.012

Error rates. Just as in the analysis of the reaction times, the interaction between proficiency, stimulus, and response direction was not significant in the error rate analysis, $F < 1$. We obtained significant effects only for the interaction between stimulus and response direction, $F(2, 372) = 12.63$, $p < .001$, $\eta_p^2 = .064$, the interaction between proficiency and response direction, $F(1, 186) = 5.30$, $p = .022$, $\eta_p^2 = .028$, and the main effect of response direction, $F(1,186) = 10.59$, $p = .001$, $\eta_p^2 = .054$. The interaction between proficiency and stimulus, $F(2, 372) = 1.15$, $p = .317$, $\eta_p^2 = .006$, the main effect of proficiency, $F(1, 186) = 3.34$, $p = .069$, $\eta_p^2 = .018$, and the main effect of stimulus, $F < 1$, were not significant.

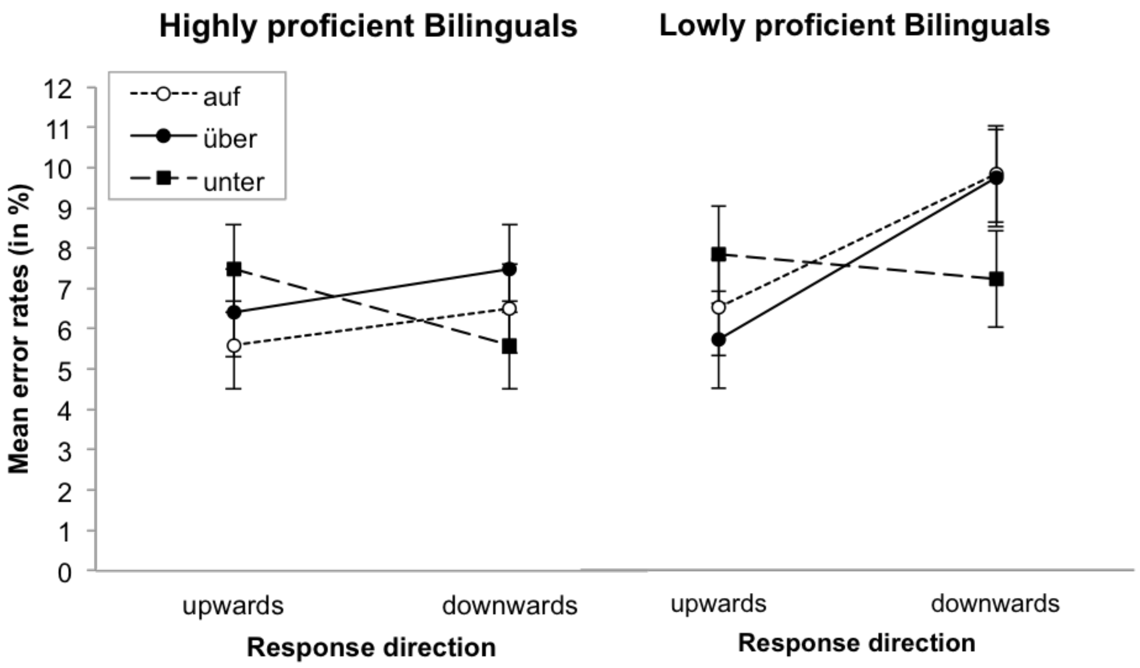


Figure 15. Mean percentage of errors of the high-proficiency and the low-proficiency group as a function of response direction and stimulus. Error bars represent 95% confidence intervals (as per Masson & Loftus, 2003).

When we looked at the high-proficiency and the low-proficiency group separately, we found a similar pattern as in the reaction time analysis: The interaction effect of stimulus and response direction was significant in both groups (high-proficiency group: $F(2, 182) = 5.99$, $p = .003$, $\eta_p^2 = .062$; low-proficiency group: $F(2, 190) = 6.85$, $p = .001$, $\eta_p^2 = .067$). While the main effect of response direction was significant for the low-proficiency group but not the high-proficiency group (high-proficiency group: $F < 1$; low-proficiency group: $F(1, 95) = 12.23$, $p = .001$, $\eta_p^2 = .114$), the main effect of stimulus was not significant in either of the groups (high-proficiency group: $F < 1$; low-proficiency group: $F < 1$).

While we found significant differences between compatible and incompatible responses for the words *auf* and *über*, but not for *unter* in the low-proficiency group, in the high-proficiency group the differences between compatible and incompatible responses were significant for the words *über*, and *unter*, but not for *auf* (see Table 9).

Table 9: *Comparisons of the error rates in compatible and incompatible trials per proficiency group and word*

		df	*t*	*p*
High-Proficiency Group				
	auf	91	-0.99	.326
	über	91	-2.78	.007
	unter	91	2.08	.040
Low-Proficiency Group				
	auf	95	-3.11	.002
	über	95	-4.05	<.001
	unter	95	0.65	.519

3 General Discussion

The aim of the current thesis was to investigate the theory of experiential traces within the framework of embodiment. This theory states that interactions with the world leave experiential traces in the brain, which are later reactivated when a person encounters linguistic stimuli. Thus, language comprehension can be understood as a reactivation of those experiential traces.

In three studies, we focused on single word processing and investigated the processing of nouns, verbs, and spatial prepositions. In a first step, in Study 1, we investigated the intriguing results obtained in different behavioural studies (e.g., Ahlberg et al., 2013; Marino et al., 2011) when investigating effector-specific nouns and action verbs. Next, we investigated whether the experiential traces account can also be extended to second language processing, see Study 2 and Study 3. Although a lot of research on embodiment has been conducted with a focus on first language processing, little is known about whether these theories can be extended to second language processing as well.

I will begin this discussion by briefly summarizing the main findings of the three studies before going deeper into explaining the limitations as well as the implications of the three studies individually and this dissertation project in total.

3.1 Summary of main findings

Study 1 investigated the processing of nouns referring to concrete entities and compared it with the processing of action words in three experiments. The aim of the study was to shed more light on the reasons for the processing differences between nouns and verbs. Therefore, it focused on temporal characteristics of the reading process as well as on characteristics concerning levels of processing. Study 1 confirmed the processing differences between effector-specific nouns and action verbs found by Ahlberg et al. (2013) by replicating its results. Interestingly, processing differences for nouns and verbs were found to be dependent on the task. The effector-specific information encoded in the nouns was accessed automatically and thereby facilitated (in compatible trials), or interfered with (in incompatible trials) the movement response that needed to be executed, and this was true for all given tasks. In contrast, the effector-specific information in verbs was not automatically activated, meaning the facilitation or interference caused by the word was task-dependent. In other words, the compatibility effect was only found in a task in which active processing was required. This suggests that the information encoded in verbs and nouns referring to concrete entities are accessed differently. The findings suggest that in nouns, the meaning is more obvious and less arbitrary, meaning that it can be more easily accessed and automatically activated. On the contrary, it seems that verb meaning is more complex and needs deeper processing to reactivate an experiential trace related to effector specificity.

While Study 2 and Study 3 aimed at extending the evidence for the reactivation of experiential traces to spatial prepositions, another goal was to investigate whether the theory of experiential traces is also applicable to L2 processing. Additionally, different factors were investigated that are thought to have an influence on the building of experiential traces in an L2. More specifically, Studies 2 and 3 focused on the influence of the L1 on L2 processing, as well as on the role of language proficiency. Furthermore, the impact of the age of acquisition of the L2 was explored. Therefore, Study 2 and 3 investigated the processing of spatial prepositions in L1 and L2 speakers of German in two different populations: adults and children, respectively. While both studies found evidence for the activation of experiential traces in L1 as well as in L2 speakers of German, differences were found depending on the age of acquisition of the L2. More precisely, for early/children learners of an L2, no processing differences compared to L1 speakers of German were obtained. In contrast, for late/adult learners of German as L2, processing differences were found. These results suggest that in adult learners of German as L2, the experiential traces connected to the L2 rely on the experiential traces of the L1. Additionally, we found some evidence that language proficiency is likely to moderate this effect. The results indicate that the higher the language proficiency, the lesser the influence of the L1 on the processing of the L2. In sum, these results suggest that an L2 can be embodied in a similar way as the L1. Moreover, this project provides evidence for the reactivation of experiential traces not only in L1, but also L2 processing. Furthermore, the results of this dissertation project provide fruitful insights into L1 and L2 processing with implications for embodied learning strategies. In addition, this project lays the groundwork for potential further studies, which are discussed below.

3.2 Reactivation of experiential traces in different word classes

According to the experiential traces account (Zwaan & Madden, 2005), every interaction with our environment leaves a trace of experience in our brain. Through these traces, the different features of objects and the words they refer to are combined. When the same words are encountered again, these experiential traces are activated and promote comprehension.

Action-related compatibility effects are commonly interpreted as evidence for the embodied cognition account. Evidence for such compatibility effects have been reported in studies investigating sentences (e.g., Glenberg & Kaschak, 2002; Olmstead et al., 2009; Scorolli & Borghi, 2007) as well as individual words (e.g., Dudschig et al., 2012; Lachmair et al., 2011; Marino et al., 2011; Mirabella et al., 2012). Thus, the underlying reasoning is that processing words or sentences referring to particular actions leads to an activation of experiential traces stemming from performing the respective actions in the past. This in turn explains why after reading linguistic material describing performing a particular action, matching or mismatching action is facilitated or hindered, respectively.

However, results are mixed with regard to verbs describing actions that are performed with a specific effector. The processing differences in verbs and nouns obtained in Study 1 suggest that the information encoded in an experiential trace is not reactivated similarly for all word classes. In general, neuropsychological studies supporting the embodiment account often strongly focus on action verbs to show the similarities in brain activation between read and executed actions (e.g., Pulvermüller et al., 2001; Hauk et al., 2004). However, Ahlberg et al. (2013) found effector-specific compatibility effects for effector-related nouns but not for effector-related action verbs in a behavioural study. More specifically, participants in this experiment responded to the font colour of an effector-related word (e.g., *cup*, *kicking*) by pressing a button with their hand or their foot. Responses in compatible trials (e.g., hand response after reading a hand-related word) were faster than responses in incompatible trials (e.g., foot response after reading a hand-related word). However, this was true only for nouns (e.g., *cup*), not for verbs (e.g., *grasp*).

In Study 1, we were able to replicate this finding. Nonetheless, we also found that the reactivation of the effector-specific information encoded in action verbs did take place when the words were processed in a lexical decision task. These results suggest a difference in the processing of nouns and verbs. More precisely, nouns seem to automatically activate effector-specific information. In contrast, in the case of verbs, participants need to be forced to access their mental lexicon before evidence can be found for effector-related activation.

This result is also in line with recent findings in neuropsychological studies suggesting that the degree of motor activation in the premotor areas elicited by action verbs might also be sensitive to attentional and situational factors (for an overview see Kemmerer, 2015).

However, it is of interest to ascertain what factors are responsible for the processing differences in nouns and verbs, as many differences between nouns and verbs could be responsible for the observed differences. For instance, nouns are mostly learned before verbs (Gentner, 1982) and might have more experiences connected to them than verbs. Therefore, it seems likely that experiential traces are accessed more automatically for nouns than for verbs.

Neuroscience also offers a plausible explanatory framework for understanding the different processes of verbs and nouns with regard to effector-specific information. More precisely, evidence has been found that different brain regions are involved in the processing of nouns and verbs (Damasio & Tranel, 1993; Preissl, Pulvermüller, Lutzenberger, & Birbaumer, 1995; Gleichgerrcht et al., 2016). Another reason might lie in the differences in broadness of meaning between verbs and nouns. As verbs cover broader meaning than nouns (Gentner, 1981), effector-specific information might get lost in the shuffle of features that are relevant to the meaning of an effector-related action verb, at least in a task that does not require active reading for completion, as in the Stroop-like task employed in Study 1. Further studies are needed to investigate whether any of these theories or other differences between nouns and verbs can account for the observed differences in task dependency.

Although we would expect these differences to account for all the features encoded in action verbs, they seem to especially affect the reactivation of effector-specific information, as the reactivation of motion information in verbs, for instance, seems to be automatic. Evidence comes from a study of motion verbs by Dudschig et al. (2012). In their study, Dudschig et al. (2012) were able to show compatibility effects between motion verbs referring to an upward or downward motion (e.g., *rise/fall*) and upward or downward hand movements in a Stroop-like task similar to the tasks used in the studies of this dissertation project. In all their experiments, responses were faster when the word's immanent motion was compatible with the required response (e.g., upward response on *rise*). While the motion information encoded in a motion verb is part of its meaning and might be accessed directly, the effector-specific information is not necessarily part of the meaning and thus depends on the person's personal experiences. Therefore, it seems plausible that the effector-specific information encoded in verbs is only assessed indirectly. This directly leads to the question of why effector-specific information is processed differently in nouns and verbs.

Further studies are needed to investigate the differences between nouns and verbs. In addition, it might be useful to further assess which features encoded in verbs and nouns get reactivated automatically and which need active processing. Furthermore, it might also be of interest to compare different sorts of verbs. For instance, a difference might be found between transitive and intransitive verbs. Transitive verbs require one or more objects (e.g., to give), while intransitive verbs do not (e.g., to run). Therefore it seems more likely that intransitive verbs show a reactivation of experiential traces compared to transitive verbs, at least in studies on single word processing where information about the object is lacking. In addition, a difference might also be obtained for highly imaginable in contrast to low imaginable verbs.

Highly imaginable verbs might be more likely reactivate experiential automatically, while low imaginable verbs might need active processing.

In Study 2 and Study 3, we found that the spatial information encoded in prepositions also can cause compatibility effects. Taken together, we thus found evidence for the automatic reactivation of experiential traces in nouns and prepositions. Our findings suggest that while the information encoded in nouns and prepositions seems to be accessed automatically, conscious processing is needed for verbs, as not all encoded information is accessed automatically.

3.3 Flexibility of experiential traces connections

Previous research has shown that experiential traces are based on people's personal experiences with their environment. For instance, right-handers showed a preferential activation of the left premotor cortex, while left-handers showed a preferential activation of the right premotor cortex, when they needed to respond to manual-action verbs (e.g., *to throw*) versus non-manual actions (e.g., *to kneel*) in an fMRI study (Willems, Hagoort, & Casasanto, 2009). Similar effects were found for athletes compared to non-athletes. Holt and Beilock (2006) investigated whether football players and ice hockey players showed compatibility effects in a picture-naming task in contrast to novices. When judging whether a presented target had been mentioned in the previous sentence, the athletes showed quicker responses to targets associated with their respective sport-specific scenarios. Similar effects were obtained for pianists versus non-musicians in a sentence sensibility judgement task (Wolter, de la Vega, Dudschig, & Kaup, 2014, as cited in Kaup, de la Vega, Strozyk, & Dudschig, 2016). While the non-musicians showed no difference between left and right hand responses while judging sentences referring to a high or low pitch, the musicians showed compatibility effects based on their personal experience. They responded faster with the left hand to high pitch related sentences and with the right hand to low pitch related sentences (Wolter et al., 2014). These studies support the assumption that experiential traces are based on a person's individual experiences. Therefore, different individuals have different experiential traces. As learning is a lifelong process in which we gain new experiences on an everyday basis, new experiential traces are built constantly and already existing traces are extended. Furthermore, it is also likely that already existing connections might be restructured or weakened. In Studies 2 and 3 of this dissertation project, we found first evidence for such a restructuring and weakening of experiential traces. In Studies 2 and 3, we focused on the processing of spatial prepositions in L1 and L2 learners of German. When looking at the L1 learners only, differences were obtained for adults (Study 2) and children (Study 3). While the adults showed a compatibility effect for the word *über*, but not for *auf*, this was the opposite for the children, where the compatibility was found for *auf*, but not for *über*.

The findings suggest that a restructuring happens over the life course that leads to the reversal of the effect in the adults. This idea points to the flexibility of experiential traces. It is therefore possible that the connections between experiential traces can also be weakened or restructured. For instance, when it comes to learning prepositions in German, *auf* is generally learned faster than *über* due to its higher frequency in the learner's input (Bryant, 2012) as well as in total use (e.g., Quasthoff et al., 2011). Therefore, the experiential trace connections between the word *auf* and its spatial meaning are likely to be built early in life due to the possibly high frequency of co-occurrence. However, since *auf* is also frequently used in various contexts in which the spatial meaning is no longer obvious (e.g., *aufhören*, *aufmachen*, *sich auf etwas freuen* etc.), it is likely that the spatial meaning fades away over the course of one's lifetime. More precisely, as later in life new learning experiences are made with non-spatial meanings of *auf,* supplementary traces between different non-spatial meanings of *auf* and the word itself can be built. Accordingly, as the non-spatial use of *auf* becomes more prominent, the corresponding experiential traces might also become stronger, as they are used and activated more often. In turn, the spatial use might become less frequent, causing this connection to weaken. Then, when the word *auf* is encountered in the future, the strongest connections are likely to be reactivated automatically, which in this case would be the non-spatial meaning. This mechanism of restructuring through usage frequency could account for the differences we found between children and adult German speakers. Moreover, this mechanism would also be in line with the Hebbian rule (following Hebb, 1949), which represents the foundation of the building of experiential traces (Zwaan & Madden, 2005).

In sum, the results indicate that experiential traces are flexible and can be restructured depending on new learning experiences. Surprisingly, however, we still found this effect among children with an age range of eleven to fifteen years. These children have already had more than 10 years of German language input, but the restructuring seems to be not fully completed. It would be a question for future research when exactly and under which circumstances the restructuring and thereby changes in the connections between experiential traces occur. Knowing more about these mechanisms and their moderating factors could also be helpful for language learning. It could give insights into which factors foster and hinder the restructuring of experiential traces.

3.4 Embodiment in second language processing

The aim of Studies 2 and 3 was to investigate the reactivation of experiential traces regarding the spatial information encoded in spatial prepositions. While the results supported the view that the spatial information encoded in prepositions can also be activated and lead to language motor interactions, a second question was also addressed in these two studies. Studies 2 and 3 aimed at investigating whether the theory of experiential traces could be extended to L2 processing. The results of Study 2 as well as Study 3 showed action-word meaning compatibility effects in a Stroop-like task for L1 and L2 speakers of German. All groups showed faster responses when the meaning of the word was compatible with the response direction (e.g., upward movement for *auf*) compared to incompatible trials (e.g., downward movement for *auf*). This finding suggests that experiential traces connected to spatial prepositions in the L2, in this case, are reactivated. Thus, these results support the assumption that the L2 is embodied in a similar way as the L1. Furthermore, this is in line with results from Dudschig et al. (2014) and De Grauwe et al. (2014). Likewise, we found that children (Study 3) reactivated experiential traces connected to the spatial categories in a similar way as adults (Study 2). This is of special interest, as we were able to find these effects even in relatively young children. This shows that experiential traces are already established and connected to words, even in a reading task that does not require active reading, at the age of eleven to fifteen years.

3.4.1 Age differences in age of acquisition

Studies 2 and 3 investigated adult learners and schoolchildren respectively. Therefore, it was also possible to compare early L2 learners, children who learned German until the age of six (Study 3) with late L2 learners, adults (Study 2). Interestingly, we found that the adult learners differed from the children with respect to the influence of the L1. For the bilingual schoolchildren, who were mainly simultaneous bilinguals, we found no processing differences between monolingual and bilingual speakers of German. In contrast, Study 2 showed differences for the adult L2 speakers depending on their native language. Our data suggest that bilingual children who learn German rather early in life do not differ from monolingual German children with regard to the processing of German prepositions. This was even the case when their L1 differed from German in terms of spatial categorization. Although different studies have shown that the acquisition of spatial prepositions leads to particularly persistent difficulties (Alonso et al., 2016; Bryant, 2012; Coventry & Garrod, 2004; Coventry et al., 2012; Grießhaber, 1999; Ijaz, 1986; Lütke, 2008; Munnich & Landau, 2010), if language acquisition happens early enough in life, a native-like acquisition seems to be possible. The acquisition of the German prepositions *auf* and *über* was found to happen before the age of five years, with *auf* being acquired at the age of 2.9 years and *über* at the age of 4.4 years on average (Grimm, 1975). Nearly 70% of the bilingual schoolchildren investigated were categorized as simultaneous bilinguals with an age of acquisition before the age of three (Rothweiler & Kroffke, 2006). Yet another 10% had acquired German as L2 between the ages of three and five and were categorized as early-successive bilinguals (Rothweiler & Kroffke, 2006). Therefore, it could be assumed that a native-like acquisition of German prepositions might be possible when the L2 is acquired before or around the same age as the L1 acquisition of prepositions takes place. However, different acquisition scenarios are imaginable for the simultaneous bilinguals and the early-successive bilinguals. It seems more likely that simultaneous bilinguals parallelly develop two separate systems of experiential traces, while early-successive bilinguals might rather show an L1 influence on the building of L2 connected experiential traces. Therefore, further research is needed to investigate the two groups of bilinguals separately in more detail.

Nevertheless, taken together, our findings support the view that early bilingual children learn German in a similar way as monolingual children learn German as L1. More precisely, in both cases the children seem to be able to build similar experiential traces as children who learned German as L1, even when they acquire German during their time in pre-school, as was also described by Tracy and Gawlitzek-Maiwald (2000) as well as Rösch (2011).

Furthermore, the differences found in the adult L2 speakers actually depended on their L1. In line with Slobin (1996), this suggests that in late L2 learning, the L1 categorization has a major influence on the L2 categorization. It seems to be quite difficult to restructure the learned categorization of the L1 later to fit the categorization of the L2. This can be inferred from the finding that we still found an influence of the L1 in the processing of the L2 for our participants, who were quite proficient already. The role of language proficiency will be discussed in the following section in more detail.

Taken together, we provide additional evidence for the reactivation of experiential traces connected to prepositions for children and adults with German as L1 or L2. Our results suggest that an L2 can be learned and processed in a similar way as the L1 when learned early in life, which is also in line with Johnson & Newport (1989). However, when an L2 is acquired later in life, as an adolescent or adult, the L1 plays a major role in the processing of the L2. However, our results only investigated the processing of German as L1 and L2, and it is not possible to draw conclusions with respect to languages other than German. Future studies would need to investigate other languages, as they might differ from German with respect to grammar and therefore the L2 learning process.

3.4.2 The influence of language proficiency

Another factor that could be relevant in the reactivation of experiential traces of an L2 is language proficiency. More specifically, it is possible that language proficiency moderates the influence of the L1 on the L2. For instance, with increasing language proficiency, a decrease in the influence of the L1 on the L2 is likely (Vukovic, 2013). In addition, although it remains unclear how L1 and L2 words are represented, in one bilingual lexicon or rather in two distinct lexicons (Dijkstra & Van Heuven, 2002; Kroll & Stewart, 1994), in the RHM (Kroll & Stewart, 1994) as well as the BIA+ Model (Dijkstra & Van Heuven, 2002), language proficiency is seen as a moderating factor. While in the RHM, language proficiency strengthens the direct connections between L2 words and their concepts in addition to the links between L1 and L2 representations, in the BIA+ Model, it is seen as promoting the selection process. Similar assumptions hold for the coactivation account (Kroll et al., 2014; Blumenfeld, & Marian, 2013), which assumes that bilinguals always activate both languages. Cognitive control is needed to supress the language representations that are currently not needed. For instance, the higher the language proficiency in the L2, the easier it is possible to supress the L1. In addition, it also leads to better control of the L2. In other words, language proficiency is thought to influence whether the L1 or L2 is more dominant, but at the same time it also determines the controllability of the two languages.

In Study 2 and Study 3, we found support for the view that language proficiency might indeed have a moderating influence on experiential traces and the influence of the L1 on L2 processing. In Study 2, differences were found between mid-to-low proficient and highly proficient L2 speakers. For the mid-to-low proficiency group, *auf* showed a significant compatibility effect, while for *über* no significant compatibility effect was obtained. In contrast, in the high proficiency group, both *auf* and *über* showed significant compatibility effects. This effect was measured for the group of L2 speakers whose L1 has a different spatial categorization than German (e.g., Turkish), as it does not make further distinctions within the upper subspace. These results are in line with the assumption that *über* is acquired later than *auf* in L2 acquisition, a similar learning order to the one assumed for German L1 learners (Bryant, 2012). It suggests that *auf* is recognized as an equivalent to the term describing the upper subspace in the L1, which can be learned faster. Later, a restructuring is needed to incorporate *über* as well. This is seen, for example, in the reactivation of experiential traces in highly-proficient learners of German as L2.

However, the sample sizes in Study 2 were rather small after subcategorization. Therefore, it is not possible to draw stable inferences from these analyses. Moreover, language proficiency was measured via a subjective measure, asking the participants for a personal evaluation according to the CEFR (Verhelst et al., 2009). Therefore, in Study 3, we included an objective measure in the form of the C-Test (Grotjahn, 1992, 2014).

The results of Study 3 also gave indications of a language proficiency influence, although the effects were only marginal. Descriptively, the children's results were similar to the adult results in Study 2 with regard to language proficiency differences. However, these differences were not significant in the analyses. Moreover, around 70% of the participants were simultaneous bilinguals who learned the L2 before the age of three, and therefore they were also quite proficient. Due to this high proficiency, it is possible that the children were able to suppress their L2 sufficiently to not affect the processing of the German prepositions.

As the sample of late learners of German as L2, children who learned German as L2 after the age of three, in Study 3 was rather small, we were not able to compare them to the group of early bilinguals in the same study. In addition, language proficiency was quite high overall which may have lead to smaller differences between the two proficiency groups. For instance, the low-proficiency bilinguals still showed a mean proficiency score of around 70%. Nevertheless, this group was also more heterogeneous with respect to language proficiency than the high-proficiency group.

In future studies, it would be interesting to look at these factors in more detail by systematically testing and directly comparing performance in a sample of children with childlike L2 acquisition vs. simultaneous or early-successive bilinguals with low vs. high proficiency.

3.5 Implications for embodied learning

Taken together, the results imply that the early acquisition process of German as L2 differs from the late acquisition process of German as L2 in adults, which is in line with research on L2 acquisition in general (e.g., Birdsong, 2006; Hyltenstam & Abrahamson, 2003; Johnson & Newport, 1989; Tracy & Gawlitzek-Maiwald, 2000). When a L2 is learned at the same time as the L1, the same learning mechanisms can be used. Therefore, it is likely that the same mechanism of co-occurrences and experiences with one's surroundings supports experiential trace building in L1 and L2. Throughout the life course, these experiential traces can be modified, extended, reduced, strengthened, or weakened.

Nevertheless, when adults learn a second language, the co-occurrences and experiences with the L2 are not as frequent and as rich as when a child learns a L2. Furthermore, the L1 can be assumed to be already strongly consolidated, which makes a restructuring or the building of experiential traces for an L2 more difficult. It is nearly impossible to learn the L2 without the L1, as stated by Slobin (1996) and Lucy (2011). Accordingly, the L1 influences our perception of our environment and guides a person's awareness in learning an L2. Additionally, in the coactivation account (Blumenfeld & Marian, 2013; Kroll et al., 2014), it is assumed that both the L1 and L2 are similarly activated. Therefore, in the coactivation account, L1 and L2 are seen to influence each other. As a result, a certain level of language proficiency would be needed to control both languages and suppress the not-needed language.

The results of the present dissertation project support the idea that the experiential traces of the L1 are first extended to the L2, but then need to be restructured later for concepts and categories in which the L2 does not fit the L1. Although this assumption needs further testing, it provides a possible approach on how to improve the learning of an L2 later in life.

Designing the learning process as more experiential in nature might promote the building of experiential traces connected to the L2 and thereby facilitate the learning of the L2 in general. By enriching the learning input with co-occurrences of L2 words and multimodal direct experiences, the building of experiential traces might be fostered. Moreover, the learning process would thereby be designed to be more similar to the process of L1 learning. It might also help people learn categorizations that do not exist in the first language by providing experiences with the L2. Thereby, experiential traces connected to the L2 could be built faster.

For instance, a study by Nakatsukasa (2016) on the efficacy of gestures on the acquisition of locative prepositions showed that gestures in combination with recast enhanced the learning of locative prepositions in a delayed post-test, in contrast to the recast only condition. This use of gestures can be seen as one form of embodied learning, and it would be interesting to further investigate the impact of these sorts of embodied learning strategies on L2 acquisition in future intervention studies. In foreign language learning and especially the acquisition of foreign language words, gestures and actions are widely used to support learning (for a review see Macedonia & Kriegstein, 2012). More precisely, action words or phrases are memorized better and for longer when learning is accompanied by enactment (e.g., Engelkamp & Krumnacker, 1980; Mecklenbräuker, Steffens, Jelenec, & Goergens, 2011; Saltz & Donnenwerth-Nolan, 1981). The possible mechanisms underlying the positive effects of gestures on learning include the grounding in one's own body as well as the multi-modality of the learning experience (Macedonia & Kriegstein, 2012). Other examples of the positive effect of embodied learning techniques come from other research areas such as number processing. In this research, the enactment of a mental number line or using a digital dance mat were found to improve the spatial representation of numerical magnitudes and thereby mathematical abilities (Fischer et al., 2016; Link, Moeller, Huber, Fischer, & Nuerk, 2013).

Taken together, future research needs to investigate the effect of embodied learning techniques on L2 learning in general and whether they might not only foster vocabulary learning but also the acquisition of grammatical rules and categorization differences between L1 and L2 in particular.

3.6 Discussing functional relevance in embodied cognition

Although the evidence for embodied cognition theories of language processing is steadily growing in the form of interaction effects between language processing and motor actions, the functional relevance of these interactions remains an open question. A large number of behavioural studies have found that language processing has an influence on executed motor actions. Additionally, neuropsychological studies have found motor cortex activation occurring during language processing. However, there is still a controversial debate on to what extent these language motor interactions are actually necessary for language comprehension, also referred to as the necessity question (Fischer & Zwaan, 2008; van Elk, Slors, & Bekkering, 2010), or whether they can be seen as an epiphenomenon.

The results of the present dissertation project, especially the results of Studies 2 and 3, suggests that experiential traces are flexible and built during learning processes. The obtained differences between early and late learners of German as L2 might have implications for the further development of embodied learning techniques. Reported studies on positive embodiment effects on learning and memorizing in vocabulary learning (for a review see Macedonia & Kriegstein, 2012) and number processing (Fischer et al., 2016; Link et al., 2013) make it seem highly unlikely that language motor interactions have no functional role at all. However, when reviewing studies with implications on the functional relevance debate, partially contradictory results were obtained. In neuroscience, for instance, the comprehension of action words was found to be impaired selectively in patients with Parkinson (Boulenger et al., 2008) and apraxia (Buxbaum & Saffran, 2002). This suggests that sensory motor simulations are needed to comprehend action words, meaning that these simulations can hardly be seen as only a by-product of language processing (Horchak, Giger, Cabral, & Pochwatko, 2014). In addition, in studies using TMS (e.g., Pulvermüller, Hauk, Nikulin, &

Ilmoniemi, 2005; Willems, Labruna, D'Esposito, Ivry, & Casasanto, 2011), stimulation of the premotor cortex facilitated responses to action verbs in a lexical decision task, which was interpreted as the premotor cortex having a functional role in action-language understanding. On the contrary, in different neurological studies a connection between motor neuron dysfunctioning and disrupted processing of action-related language was found (Arévalo et al, 2007; Arévalo, Baldo, & Dronkers, 2012; Kemmerer, Rudrauf, Manzel, & Tranel, 2012; see also Taylor & Zwaan, 2012), with slower or less fluent performance on action or effector-related words, while general lexical or semantic processing remained intact (Taylor & Zwaan, 2012). This finding suggests that motor activation is not necessary for minimal action language comprehension.

Behavioural studies found similar results, with impairment in language comprehension when the respective effectors were occupied with a patty cake or tapping task executed in parallel (Strozyk, Dudschig, & Kaup, 2015, as cited in Kaup et al., 2016; Yee, Chrysikou, Hoffman, & Thompson-Schill, 2013). In the study by Yee et al. (2013), it was harder for participants to comprehend words referring to manipulable objects when the manual system was occupied by a parallelly executed patty cake task. Strozyk et al. (2015) found that a hand-tapping task had larger interference on hand-related words than foot-related words (e.g., *cup* and *handbag* vs. *shoe* and *football*). Their participants conducted a lexical decision task and responded slower to hand-related words than foot-related words when a hand-tapping task was executed in parallel. In addition, Shebani and Pulvermüller (2013) found an effect of a tapping task on memory for action words. Hand tapping interfered more with the memory of arm-related words than leg-related words, while foot tapping interfered more with the memory of leg-related words than arm-related words.

Although these studies support the view that sensory motor activation has a functional role in language processing, sensory motor activation does not seem to be compulsory to fulfil the tasks at hand. Despite the occupation of the motor system, these tasks were all still executable. This suggests that sensory motor activation fosters language comprehension, but language comprehension might also be possible without this activation. However, it is also important to note that in order to fulfil both tasks, both the tapping and the language processing might need to be equally slowed down. In the occupation studies, it was assumed that the hand tapping or patty cake task occupied the respective motor region completely. However, the motor area's working capacity might also be split evenly between the two tasks. Thus, it is possible that not only language processing but also for instance the tapping task was slowed down, making it possible to access the motor area for language processing in a limited way. However, a study by Postle, Ashton, McFarland, and de Zubicaray (2013) investigated the effects of a reading task on a hand-tapping task. Postle et al. (2013) found general differences between the base line tapping rate and the tapping rate while reading words or sentences related to different body parts out loud. However, the reduction in tapping rate did not differ with respect to the related body part. Reading hand-related words or sentences did not lead to stronger impairment of hand tapping than words or sentences related to other body parts. Although this suggests that in this dual task paradigm the tapping rate can also be reduced by the parallel hand-tapping task, it was not specifically impaired by hand-related words. Rather, it was generally impaired by a subsequent motor task, namely reading out loud. Future studies need to investigate this further. Studies that assess a baseline for either of these tasks, the language processing and the occupation task, are particularly needed. Thereby, it would be possible to compare performance on dual tasks with performance on each of the single tasks directly, allowing inferences on the distribution of capacity to be drawn.

On the basis of the studies described above, one could assume that sensory motor activation is not necessary for action understanding, but rather functions to enrich representations in order to support language comprehension and action perception (Taylor & Zwaan, 2012). As a result, sensory motor activation might lead to more fluent information processing (Fischer & Zwaan, 2008). This also opens a discussion about the degree of embodiment in language comprehension (Chatterjee, 2010). Actually, three different versions can be differentiated that differ with regard to the degree of sensory motor involvement in language processing (Kaup et al., 2016).

The view of strong embodiment is also named the one-format view. In this account, language comprehension is not possible without sensory motor activation. As a result of the evidence reported above, this view has been called into question, as it presumes that sensory motor activation is necessary for language comprehension. In contrast, the weakest view of embodiment, also called the word-based resonance view, sees sensory motor activation only as a by-product of language comprehension, which is thereby epiphenomenal. Studies of patients with brain injuries could contribute to disproving this view, as they have found functional impairment of motor language comprehension (Boulenger et al., 2008; Buxbaum & Saffran, 2002). A third view of embodiment is subsumed under dual-format views. These views represent a moderate view of embodiment in which basic language comprehension might be possible without sensory motor activation, but sensory motor activation is seen as enriching the mental representation. In this view, language comprehension would also be possible without, for instance, explicit expert knowledge of playing basketball when reading about it. However, the more experience the reader has with the topic described, the more sensory motor activation would occur to enrich the representation and thereby, for instance, a mental model of the situation described in the text (Kaup et al. 2016; Taylor & Zwaan, 2012).

As of yet, there is not enough evidence to fully reject one of these views. More research is needed to investigate further whether sensory motor activation is sufficient for language comprehension and under which circumstances embodiment supports language comprehension. This is of special interest with regard to possible interventions or language teaching programs that are based on the embodiment framework to enhance language learning.

4 Conclusion

In this dissertation project, the experiential traces account within the embodiment framework was investigated. A special focus lay on single word processing of different word classes, namely nouns, verbs, and spatial prepositions. Furthermore, the present work intended to investigate the experiential traces account not only in terms of L1 processing, but also whether this account can be extended to L2 processing.

The reported studies provide evidence for the embodiment account, supporting the view that readers activate experiential traces when reading words or sentences that stem from prior interactions with the referents of the linguistic expressions. These results cannot be explained by an amodal account of language comprehension. The obtained language-motor compatibility effects suggest that experiential traces are activated (1) in different word classes as nouns, verbs, as well as spatial prepositions and (2) in L1 and L2 processing. Our results, however, also show differences between effector-related nouns and effector-related action verbs, suggesting that not all information encoded in an experiential trace is reactivated in the same way in all word classes. Whereas for nouns, effector-specific compatibility effects were automatic and task independent, the same effects for action verbs were task dependent and therefore were only present in a task that required accessing the mental lexicon. Furthermore, we were able to find evidence for the reactivation of experiential traces in L2 processing for different learner groups with different native languages as well as differences in the age of acquisition. In addition, we found differences between early and late L2 learners, suggesting that early L2 learners are able to acquire the L2 in a similar way as the L1, while the influence of the L1 plays an important role in late learning of an L2.

This project provides a good starting point for further investigations of the underlying mechanisms in the acquisition of an L2 within the embodiment framework, especially with a focus on the factors age of acquisition and language proficiency. Another key aspect that should be investigated is the impact of embodied learning strategies on late L2 learners. These learners seem to encounter the greatest difficulties in restructuring the experiential traces connected to the learned L1 in order to build the experiential traces for the L2.

References

Ahlberg, D. K., Dudschig, C., & Kaup, B. (2013). Effector specific response activation during word processing. In M. Knauff, M. Pauen, N. Sebanz & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 133-138). Austin, TX: Cognitive Science Society.

Ahlberg, D. K., Bischoff, H., Strozyk, J. V., Bryant, D., & Kaup, B. (2016). *Grounded cognition: Comparing language-space interactions in L1 and L2.* Manuscript submitted for publication.

Ahlberg, D. K., Bischoff, H., Strozyk, J. V., Bryant, D., & Kaup, B. (2016). *How do German bilingual schoolchildren process German prepositions? – A study on language-motor interactions.* Manuscript submitted for publication.

Ahlberg, D.K., Strozyk, J.V., Dudschig, C., & Kaup, B. (2016). *Processing differences of effector-related nouns and verbs: Discussing effector-specific compatibility effects.* Manuscript submitted for publication.

Aksan, Y., Aksan, M., Koltuksuz, A., Sezer, T., Mersinli, Ü., Demirhan, U. U.,…Yıldız, İ. (2012). Construction of the Turkish National Corpus (TNC). Proceedings of the eight international conference on language resources and evaluations, Istanbul, Turkey. Retrieved from http://www.lrec-conf.org/proceedings/lrec2012/pdf/ 991_Paper.pdf

Alonso, R., Cadierno, T., & Jarvis, S. (2016). Cross-linguistic influence in the acquisition of spatial prepositions in English as a L2. In R. Alonso (Ed.), *Cross-Linguistic Influence in SLA* (pp. 93-120). Bristol/Buffalo/Toronto: Multilingual Matters.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Ansorge, U., Kiefer, M., Khalid, S., Grassl, S.I., & Koenig, P. (2010). Testing the theory of embodied cognition with subliminal words. *Cognition*, *116*, 303-320. doi: 10.1016/j.cognition.2010.05.010

Arévalo, A., Perani, D., Cappa, S. F., Butler, A., Bates, E., & Dronkers, N. (2007). Action and object processing in aphasia: From nouns and verbs to the effect of manipulability. *Brain and Language*, *100*, 79–94. doi: 10.1016/j.bandl.2006.06.012

Arévalo, A. L., Baldo, J. V., & Dronkers, N. F. (2012). What do brain lesions tell us about theories of embodied semantics and the human mirror neuron system? *Cortex*, *48*, 242-254. doi: 10.1016/j.cortex.2010.06.001

Aristotle (1961) *De anima, Books II and III,* (D. W. Hamlyn. Trans.) Oxford University Press. (Original work published in 4th century B.C.)

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology*, *16*, 1-6. doi: 10.1016/j.cub.2006.07.060

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577-660. doi: 10.1017/S0140525X99002149

Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, *18*, 513-562. doi: 10.1080/01690960344000026

Barber, H. A., & Kutas, M. (2007). Interplay between computational models and cognitive electrophysiology in visual word recognition. *Brain Research Reviews*, *53*, 98-123. doi: 10.1016/j.brainresrev.2006.07.002

Baur, R., Chlosta, C., & Goggin, M. (2011). Der Cornelsen C-Test. In *Fördermaterialien Deutsch als Zweitsprache 5/6* (pp. 3-22). Berlin: Cornelsen Verlag.

Baur, R., & Goggin, M. (2005a). C-Test Klassenstufe 6. Unpublished instrument.

Baur, R., & Goggin, M. (2005b). C-Test Klassenstufe 7. Unpublished instrument.

Becker, A., & Carroll, M. (1997). *The acquisition of spatial relations in a second language*. Amsterdam: Benjamins.

Bergen, B., Lau, T. T. C., Narayan, S., Stojanovic, D., & Wheeler, K. (2010). Body part representations in verbal semantics. *Memory & Cognition*, *38*, 969-981. doi: 10.3758/MC.38.7.969

Birdsong, D. (2006) Age and Second Language Acquisition and Processing: A Selective Overview. *Language Learning*, *56*, 1467-9922. doi: 10.1111/j.1467-9922.2006.00353.x

Blumenfeld, H. K., & Marian, V. (2013). Parallel language activation and cognitive control during spoken word recognition in bilinguals. *Journal of Cognitive Psychology*, *25*, 37-41. doi: 10.1080/20445911.2013.812093.

Borghi, A. M. (2011). Simulating the elimination of simulation: the case of language comprehension. *Journal of Mental Imagery, 35*(3-4), 27-33.

Borregine, K. L., & Kaschak, M. P. (2006). The action-sentence compatibility effect: It's all in the timing. *Cognitive Science*, *30*, 1097-1112. doi: 10.1207/s15516709cog0000_91

Boulenger, V., Hauk, O., & Pulvermüller, F. (2009). Grasping ideas with the motor system: Semantic somatotopy in idiom comprehension. *Cerebral Cortex*, *19*, 1905-1914. doi: 10.1093/cercor/bhn217

Boulenger, V., Roy, A. C., Paulignan, Y., Déprez, V., Jeannerod, M., & Nazir, T. A. (2006). Cross-talk between language processes and overt motor behavior in the first 200 ms of processing. *Journal of Cognitive Neuroscience, 18*, 1607-1615. doi: 10.1162/jocn.2006.18.10.1607

Boulenger, V., Silber, B. Y., Roy, A. C., Paulignan, Y., Jeannerod, M., & Nazir, T. A. (2008). Subliminal display of action words interferes with motor planning: A combined EEG and kinematic study. *Journal of Physiology-Paris*, *102*, 130–136. doi: 10.1016/j.jphysparis.2008.03.015

Botvinick, M. M., Braver, T.S., Barch, D.M., Carter, C.S., & Cohen, J.D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624-652. doi: 10.1037/0033-295X.108.3.624.

Bowerman, M. (1996). Learning how to structure space for language. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 385-436). Cambridge, MA: MIT Press.

Bowerman, M., & Pederson, E. (1992). Topological relations picture series. In S. C. Levinson (Ed.), *Space stimuli kit 1.2* (p. 51). Nijmegen: Max Planck Institute for Psycholinguistics.

Bryant, D. (2012). Lokalisierungsausdrücke im Erst- und Zweitspracherwerb: typologische, ontogenetische und kognitionspsychologische Überlegungen zur Sprachförderung in DaZ (Thema Sprache - Wissenschaft für den Unterricht, 2). Baltmannsweiler: Schneider Verlag Hohengehren.

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition, 13*, 359-371. doi: 10.1017/S1366728909990344

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. Brain Research, *Cognitive Brain Research*, *24*, 355–363. doi: 10.1016/j.cogbrainres.2005.02.020

Buxbaum, L. J., & Saffran, E. M. (2002). Knowledge of object manipulation and object function: Dissociations in apraxic and nonapraxic subjects. *Brain and Language*, *82*, 179-199. doi: 10.1016/S0093-934X(02)00014-7

Chatterjee, A. (2010). Disembodying cognition. *Language and Cognition*, *2*, 79-116. doi: 10.1515/langcog.2010.004

Chersi, F., Thill, S., Ziemke, T., & Borghi, A. M. (2010). Sentence processing: Linking language to motor chains. *Frontiers in Neurorobotics*, *4*, 1-9. doi: 10.3389/fnbot.2010.00004

Choi, S., & Hattrup, K. (2012). Relative contribution of perception/cognition and language on spatial categorization. *Cognitive Science*, *36*, 102-129. doi: 10.1111/j.1551-6709.2011.01201.x

Coltheart, M., Davelaar, E., Jonasson, J. T , & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 64-85). Hillsdale, NJ: Erlbaum.

Coventry, K. R., & Garrod, S. C. (2004). Saying, seeing and acting: The psychological semantics of spatial prepositions. New York, NY: Psychology Press.

Coventry, K., Guijarro-Fuentes, P., & Valdés, B. (2012). On the first and second language acquisition of spatial language. *Spatial Cognition & Computation*, *12*, 219-230. doi: 10.1080/13875868.2012.713058

Damasio, A. R., & Tranel, D. (1993). Nouns and verbs are retrieved with differently distributed neural systems. *Proceedings of the National Academy of Sciences*, *90*, 4957-4960. doi: 10.1073/pnas.90.11.4957

Davies, M. (2008) Word frequency data – Corpus of Contemporary American English. Retrieved May 4th, 2016, from http://www.wordfrequency.info/

De Grauwe, S., Willems, R. M., Rueschemeyer, S. A., Lemhöfer, K., & Schriefers, H. (2014). Embodied language in first-and second-language speakers: neural correlates of processing motor verbs. *Neuropsychologia*, *56*, 334-349. doi: 10.1016/j.neuropsychologia.2014.02.003

de Vega, M., Moreno, V., & Castillo, D. (2013). The comprehension of action-related sentences may cause interference rather than facilitation on matching actions. *Psychological Research, 77*, 20-30. doi: 10.1007/s00426-011-0356-1

de Vega, M., & Urrutia, M. (2011). Counterfactual sentences activate embodied meaning: An action-sentence compatibility effect study. *Journal of Cognitive Psychology, 23*, 962-973. doi: 10.1080/20445911.2011.590471

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*, 175-197. doi: 10.1017/S1366728902003012

Dronkers, N. F., Pinker, S., & Damasio, R., (2000). Language and the aphasias. In E. R. Kandel, J. Schwartz, & T. Jessell (Eds.), *Principles in Neural Science*, (Vol. 4, 1169–1187). New York: McGraw-Hill.

Dudschig, C., de la Vega, I., & Kaup, B. (2014). Embodiment and second-language: Automatic activation of motor responses during processing spatially associated L2 words and emotion L2 words in a vertical Stroop paradigm. *Brain & Language*, *132*, 14-21. doi: 10.1016/j.bandl.2014.02.002

Dudschig, C., de la Vega, I., & Kaup, B. (2015). What's up? Emotion-specific activation of vertical space during language processing. *Acta Psychologica*, *156*, 143-155. doi: 10.1016/j.actpsy.2014.09.015

Dudschig, C., & Kaup, B. (2016). Is it all task-specific? The role of binary responses, verbal mediation and saliency for eliciting language-space associations. *Journal of Experimental Psychology: Learning, Memory & Cognition*. doi: 10.1037/xlm0000297

Dudschig, C., Lachmair, M., de la Vega, I., De Filippis, M., & Kaup, B. (2012). Do task-irrelevant direction-associated motion verbs affect action planning? Evidence from a Stroop paradigm. *Memory & Cognition*, *40*, 1081-1094. doi: 10.1037/a0024979

Duyck, W., & Warlop, N. (2009). Translation priming between the native language and a second language. *Experimental Psychology*, *56*, 173-179. doi: 10.1027/1618-3169.56.3.173

Engelkamp, J., & Krumnacker, H. (1980). Image- and motor-processes in the retention of verbal materials. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *27*, 511-533.

Epicurus (1994). *The Epicurus reader: Selected writings and testimonia.* (B. Inwood & L. P. Gerson, Trans.) Hackett. (Original work published in the 4th century B. C.)

Estes, Z., Verges, M. & Barsalou, L. W. (2008). Head Up, Foot Down: Object Words Orient Attention to the Objects‘ Typical Location. *Psychological Science*, *19*, 93-97. doi: 10.1111/j.1467-9280.2008.02051.x

Finkbeiner, M., Forster, K., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language*, *51*, 1-22. doi: 10.1016/j.jml.2004.01.004

Fischer, U., Moeller, K., Class, F., Huber, S., Cress, U., & Nuerk, H. C. (2016). Dancing with the SNARC: Measuring spatial-numerical associations on a digital dance mat. *Canadian Journal of Experimental Psychology*. doi: 10.1037/cep0000084

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology, 61*, 825-850. doi: 10.1080/17470210701623605

Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Ganzeboom, H. B. G., de Graaf, P. M., Treiman, D. J., & de Leeuw, J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, *21*, 1 – 56. doi: 10.1016/0049-089X(92)90017-B

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory, 4*, 161-178.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Language, thought, and culture* (Vol. 2, pp. 301-334). Hillsdale, NJ: Erlbaum.

Glaser, W. R. (1992). Picture naming. *Cognition*, *42*, 61-105. doi: 10.1016/0010-0277(92)90040-O

Gleichgerrcht, E., Fridriksson, J., Rorden, C., Nesland, T., Desai, R., & Bonilha, L. (2016). Separate neural systems support representations for actions and objects during narrative speech in post-stroke aphasia. *NeuroImage: Clinical*, *10*, 140-145. doi: 10.1016/j.nicl.2015.11.013

Glenberg, A. M. (1997). What memory is for: Creating meaning in the service of action. *Behavioral and Brain Sciences*, *20*, 41-50. doi: 10.1017/S0140525X97470012

Glenberg, A. M., Schroeder, J. L., & Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & Cognition*, *26*, 651-658. doi: 10.3758/BF03211385

Glenberg, A. M., Brown, M., & Levin, J. R. (2007). Enhancing comprehension in small reading groups using a manipulation strategy. *Contemporary Educational Psychology*, *32*, 389-399. doi: 10.1016/j.cedpsych.2006.03.001

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*, 558-565. doi: 10.3758/BF03196313

Goetz, E. T., Anderson, R. C., & Schallert, D. L. (1981). The representation of sentences in memory. *Journal of Verbal Learning and Verbal Behavior*, *20*, 369-385. doi: 10.1016/S0022-5371(81)90506-5

Goggin, M. (2011) C-Test Klassenstufe 9. Unpublished instrument.

Goggin, M. (2014) C-Test Klassenstufe 8. Unpublished instrument.

Grießhaber, W. (1999). *Die relationierende Prozedur. Zu Grammatik und Pragmatik lokaler Präpositionen und ihrer Verwendung durch türkische Deutschlerner.* Münster: Waxmann.

Grimm, H. (1975). On the child's acquisition of semantic structure underlying the wordfield of prepositions. *Language and Speech*, *18*, 97-119. doi: 10.1177/002383097501800201

Grotjahn, R. (Ed.) (1992). *Der C-Test. Theoretische Grundlagen und praktische Anwendungen. Vol. 1* (Manuskripte zur Sprachlehrforschung Band 39/1). Bochum: Universitätsverlag Dr. N. Brockmeyer

Grotjahn, R. (Ed.). (2014). *Der C-Test: Aktuelle Tendenzen/ The C-Test: Current Trends.* Frankfurt am Main: Lang.

Günther, K.B. (1986). Ein Stufenmodell der Entwicklung kindlicher Lese- und Schreibstrategien. In H. Brügelmann (Ed.), *ABC und Schriftsprache. Rätsel für Kinder, Lehrer und Forscher* (pp. 32–54). Konstanz: Faure.

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun - An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods*, *47*, 930-944. doi:10.3758/s13428-014-0529-0

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*, 335-346. doi: 10.1016/0167-2789(90)90087-6

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*, 301-307. doi: 10.1016/S0896-6273(03)00838-9

Hebb, D. O. (1949). *The Organisation of Behavior: A Neuropsychological Approach.* New York: John Wiley and Sons.

Heister, J., Würzner, K. M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB - eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau*, *62*, 10-20.

Holt, L. E., & Beilock, S. L. (2006). Expertise and its embodiment: Examining the impact of sensorimotor skill expertise on the representation of action-related text. *Psychonomic Bulletin & Review*, *13*, 694-701. doi: 10.3758/BF03193983

Horchak, O. V., Giger, J. C., Cabral, M., & Pochwatko, G. (2014). From demonstration to theory in embodied language comprehension: a review. *Cognitive Systems Research, 29*, 66-85. doi: 10.1016/j.cogsys.2013.09.002

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in second language acquisition. In C. Doughty, & M.H. (Eds.), *Handbook of second language acquisition* (pp. 539-588). Oxford: Blackwell.

Ijaz, I. H. (1986). Linguistic and cognitive determinants of lexical acquisition in a second language. *Language Learning*, *36*, 401-451. doi: 10.1111/j.1467-1770.1986.tb01034.x

Illes, J., Francis, W. S., Desmond, J. E., Gabrieli, J., Glover, G., Poldrack, R., Lee, C., & Wagner, A. (1999). Convergent cortical representation of semantic processing in bilinguals. *Brain and Language*, *70*, 347–363. doi: 10.1006/brln.1999.2186

James, W. (1890). *The Principles of psychology* (Vols. 1 & 2). New York: Holt.

Jirak, D., Menz, M. M., Buccino, G., Borghi, A. M., & Binkofski, F. (2010). Grasping language – A short story on embodiment. *Consciousness and Cognition*, *19*, 711-720. doi: 10.1016/j.concog.2010.06.020

Johnson, J. & Newport, E. (1989) Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60-99. doi: 10.1016/0010-0285(89)90003-0

Kaltenbacher, E., & Klages, H. (2006). Sprachprofil und Sprachförderung bei Vorschulkindern mit Migrationshintergrund. In B. Ahrenholz (Ed.), *Kinder mit Migrationshintergrund. Spracherwerb und Fördermöglichkeiten* (pp 80-97), Freiburg: Fillibach.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition*, *94*, B79-B89. doi: 10.1016/j.cognition.2004.06.005

Kaup, B., de la Vega, I., Strozyk, J., & Dudschig, C. (2016). The role of sensorimotor processes in meaning composition. In M. H. Fischer & Y. Coello (Eds.), *Foundations of embodied cognition - Volume 2: Conceptual and interactive embodiment* (pp. 46-70). London, UK: Routledge.

Khalid, S., & Ansorge, U. (2013). The Simon effect of spatial words in eye movements: Comparison of vertical and horizontal effects and of eye and finger responses. *Vision Research*, *86*, 6-14. doi: 10.1016/j.visres.2013.04.001

Kemmerer, D. (2015). Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. *Psychonomic Bulletin & Review*, *22*, 1068-1075. doi: 10.3758/s13423-014-0784-1

Kemmerer, D., Castillo, J. G., Talavage, T., Patterson, S., & Wiley, C. (2008). Neuroanatomical distribution of five semantic components of verbs: Evidence from fMRI. *Brain and Language*, *107*, 16-43. doi: 10.1016/j.bandl.2007.09.003

Kemmerer, D., Rudrauf, D., Manzel, K, & Tranel, D. (2012). Behavioral patterns and lesion sites associated with impaired processing of lexical and conceptual knowledge of actions. *Cortex*, *48*, 826-848. doi: 10.1016/j.cortex.2010.11.001

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*, 627-633. doi: 10.3758/BRM.42.3.627

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological review*, *95*, 163-182. doi: 10.1037/0033-295X.95.2.163

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363-394. doi: 10.1037/0033-295X.85.5.363

Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., & Keenan, J. M. (1975). Comprehension and recall of text as a function of content variables. *Journal of verbal learning and verbal behavior*, *14*, 196-214. doi: 10.1016/S0022-5371(75)80065-X

Kintsch, W., & Keenan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive psychology*, *5*, 257-274. doi: 10.1016/0010-0285(73)90036-4

Klein, W. (1992). *Zweitspracherwerb* (Studienbuch Linguistik; 5th ed.). Frankfurt am Main: Anton Hain.

Klepp, A., Weissler, H., Niccolai, V., Terhalle, A., Geisler, H., Schnitzler, A., & Biermann-Ruben, K. (2014). Neuromagnetic hand and foot motor sources recruited during action verb processing. *Brain and Language, 128*, 41-52. doi: 10.1016/j.bandl.2013.12.001

Köper, M., & Schulte im Walde, S. (2016). Automatically generated norms of abstractness, arousal, imageability and valence for 350 000 German lemmas. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. (...), & S. Piperidis (Eds.), *Proceedings of the 10th Conference on Language Resources and Evaluation* (LREC). Portorož, Slovenia: European Language Resources Association.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149-174. doi: 10.1006/jmla.1994.1008

Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. *Current Directions in Psychological Science*, *23*, 159-163. doi: 10.1177/0963721414528511

Kutas M., Van Petten C., Kluender R. (2006). Psycholinguistics electrified II: 1994-2005. In M. Traxler, M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 659–724). New York: Elsevier.

Lachmair, M., Dudschig, C., De Filippis, M., de la Vega, I., & Kaup, B. (2011). Root versus roof: automatic activation of location information during word processing. *Psychonomic Bulletin & Review, 18*, 1180-1188. doi: 10.3758/s13423-011-0158-x

Link, T., Moeller, K., Huber, S., Fischer, U., & Nuerk, H. C. (2013). Walk the number line - An embodied training of numerical concepts. *Trends in Neuroscience and Education*, *2*, 74-84. doi: 10.1016/j.tine.2013.06.005

Long, M. H. (1990). Maturational constraints on language development. *Studies in second language acquisition*, *12*, 251-285. doi: 10.1017/S0272263100009165

Lotze, R. H. (1852). *Medicinische Psychologie oder Physiologie der Seele*. Leipzig: Weidmann'sche Buchhandlung.

Lucy, J. (2011). Language and cognition: The view from anthropology. In V. Cook & B. Bassetti (Eds.), *Language and Bilingual Cognition* (pp. 43-68). New York: Psychology Press.

Lütke, B. (2008). Beobachtungen zur Raumreferenz in mündlichen Erzählungen von Grundschulkindern mit Deutsch als Erst- und Zweitsprache. In B. Ahrenholz (Ed.), *Zweitspracherwerb. Diagnosen, Verlauf, Voraussetzungen* (pp. 151-170). Freiburg im Breisgau: Fillibach.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163-203. doi: 10.1037/0033-2909.109.2.163

Macedonia, M., & Kriegstein, K. (2012). Gestures enhance foreign language learning. *Biolinguistics*, *6*, 393-416.

Marley, S. C., Levin, J. R., & Glenberg, A. M. (2007). Improving Native American children's listening comprehension through concrete representations. *Contemporary Educational Psychology*, *32*, 537-550. doi: 10.1016/j.cedpsych.2007.03.003

Marino, B. F. M., Gough, P. M., Gallese, V., Riggio, L., & Buccino, G. (2011). How the motor system handles nouns: a behavioral study. *Psychological Research, 77*, 64-73. doi: 10.1007/s00426-011-0371-2

Martinez de la Vega, I. J. (2013). Body-specific influences on language processing: investigating predictions of the experiential-simulations view of language comprehension (Doctoral dissertation). Eberhard Karls Universität, Tübingen.

Masson, M. E. (2015). Toward a deeper understanding of embodiment. *Canadian Journal of Experimental Psychology, 69,* 159-164. doi: 10.1037/cep0000055

Masson, M. E., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, *57*, 203-220. doi: 10.1037/h0087426

Mecklenbräuker, S., Steffens, M. C., Jelenec, P., & Goergens, N. K. (2011). Interactive context integration in children? Evidence from an action memory study. *Journal of experimental child psychology*, *108*, 747-761. doi: 10.1016/j.jecp.2010.12.002

Meisel, J. M. (2009). Second language acquisition in early childhood. *Zeitschrift für Sprachwissenschaft*, *28*, 5-34. doi: 10.1515/ZFSW.2009.002

Mirabella, G., Iaconelli, S., Spadacenta, S., Federico, P., & Gallese, V. (2012). Processing of hand-related verbs specifically affects the planning and execution of arm reaching movements. *PLoS ONE, 7:* e35403. doi: 10.1371/journal.pone.0035403

Munnich, E., Landau, B., & Dosher, B. (2001). Spatial language and spatial representation: A cross-linguistic comparison. *Cognition, 81*, 171-208. doi: 10.1016/S0010-0277(01)00127-5

Munnich, E., & Landau, B. (2010). Developmental decline in the acquisition of spatial language. *Language Learning and Development*, *6*, 32-59. doi: 10.1080/15475440903249979

Nakatsukasa, K. (2016). Efficacy of recasts and gestures on the acquisition of locative prepositions. *Studies in Second Language Acquisition*, 1-29. doi: 10.1017/S0272263115000467

National Institute of Korean Language (2005). Research of frequency of modern Korean usage [현대 국어 사용 빈도 조사 2 (Hyeondae Gugeo Sayong Bindo Josa 2)]. Retrieved May 16, 2016, from http://korean.go.kr/front/reportData/reportDataView.do?mn_id=45&report_seq=1&pageIndex=1)

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97-113. doi: 10.1016/0028-3932(71)90067-4

Olmstead, A. J., Viswanathan, N., Aicher, K. A. & Fowler, C. A. (2009). Sentence comprehension affects the dynamics of bimanual coordination: Implications for embodied cognition. *The Quarterly Journal of Experimental Psychology*, *62*, 2409-2417. doi: 10.1080/17470210902846765

Paul, L. M., Simons, G. F., & Fennig, C. D. (2016). Ethnologue: Languages of the World, Nineteenth edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com

Papua-Neuguinea. (2016, June 23). Retrieved from http://www.auswaertiges-amt.de/DE/Aussenpolitik/Laender/Laenderinfos/01-Nodes_Uebersichtsseiten/PapuaNeuguinea_node.html

Pecher, D., & Zwaan, R. A. (2005). Introduction to grounding cognition. In D. Pecher & R. A. Zwaan (Eds.) *Grounding Cognition: The role of perception and action in memory, language, and thinking.* (pp. 1-7). Cambridge: Cambridge University Press.

Postle, N., McMahon, K. L., Ashton, R., Meredith, M., & de Zubicaray, G. I. (2008). Action word meaning representations in cytoarchitectonically defined primary and premotor cortices. *Neuroimage*, *43*, 634-644. doi: 10.1016/j.neuroimage.2008.08.006

Postle, N., Ashton, R., McFarland, K., & de Zubicaray, G. I. (2013). No specific role for the manual motor system in processing the meanings of words related to the hand. *Frontiers in Human Neuroscience, 7*: 11. doi: 10.3389/fnhum.2013.00011

Preissl, H., Pulvermüller, F., Lutzenberger, W., & Birbaumer, N. (1995). Evoked potentials distinguish between nouns and verbs. *Neuroscience Letters*, *197*, 81-83. doi: 10.1016/0304-3940(95)11892-Z

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*, 576-582. doi: 10.1038/nrn1706

Pulvermüller, F., Härle, M., & Hummel, F. (2001). Walking or talking?: Behavioral and neurophysiological correlates of action verb processing. *Brain and Language, 78,* 143-168. doi: 10.1006/brln.2000.2390

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience, 21*, 793-797. doi: 10.1111/j.1460-9568.2005.03900.x

Pylyshyn, Z. W. (1973). What the mind’s eye tells the mind’s brain: A critique of mental imagery. *Psychological Bulletin*, *80*, 1-24.

Quasthoff, U., Fiedler, S., & Hallsteinsdóttir, E. (2011). *Frequency Dictionary German: Häufigkeitswörterbuch Deutsch*. Leipzig: Leipziger Uni-Verlag.

Ratcliff, R., & McKoon, G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. *Journal of verbal learning and verbal behavior*, *17*, 403-417. doi: 10.1016/S0022-5371(78)90238-4

Richardson, D. C., Spivey, M. J., Barsalou, L. W. & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, *27*, 767-780. doi: 10.1016/S0364-0213(03)00064-8

Rösch, H. (2001). *Deutsch als Zweit- und Fremdsprache*. Berlin: Akademie Verlag.

Rothweiler, M., & Kroffke, S. (2006). Bilingualer Spracherwerb: Simultane und sukzessive Mehrsprachigkeit. In J. Siegmüller & H. Bartels (Eds.), *Leitfaden Sprache-Sprechen-Schlucken-Stimme* (pp. 44-49). München: Elsevier.

Rüschemeyer, S., Zysset, S., & Friederici, A. (2005). Native and non-native reading of sentences: An fMRI experiment. *NeuroImage*, *31*, 354-365. doi: 10.1016/j.neuroimage.2005.11.047

Rüschemeyer, S. A., Brass, M., & Friederici, A. D. (2007). Comprehending prehending: Neural correlates of processing verbs with motor stems. *Journal of Cognitive Neuroscience*, *19*, 855-865. doi: 10.1162/jocn.2007.19.5.855

Ruoff, A. (1990). *Häufigkeitswörterbuch gesprochener Sprache*. Tübingen: Niemeyer.

Saltz, E., & Donnenwerth-Nolan, S. (1981). Does motoric imagery facilitate memory for sentences? A selective interference test. *Journal of Verbal Learning and Verbal Behavior*, *20*, 322-332. doi: 10.1016/S0022-5371(81)90472-2

Scorolli, C., & Borghi, A. M. (2007). Sentence comprehension and action: Effector specific modulation of the motor system. *Brain Research, 1130*, 119-124. doi: 10.1016/j.brainres.2006.10.033

Šetić, M. & Domijan, D. (2007). The influence of vertical spatial orientation on property verification. *Language and Cognitive Processes*, *22*, 297-312. doi: 10.1080/01690960600732430

Sharoff, S., Umanskaya, E., & Wilson, J. (2013). *A frequency dictionary of Russian: Core vocabulary for learners.* London, UK: Routledge.

Shebani, Z., & Pulvermüller, F. (2013). Moving the hands and feet specifically impairs working memory for arm-and leg-related action words. *Cortex, 49*, 222-231. doi: 10.1016/j.cortex.2011.10.005

Slobin, D. (1996). From ‘thought and language’ to ‘thinking for speaking’. In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70-96). Cambridge: Cambridge University Press.

Snodgrass, J. G. (1984). Concepts and their surface representations. *Journal of Verbal Learning and Verbal Behavior*, *23*, 3-22. doi: 10.1016/S0022-5371(84)90479-1

Stringer, D. (2010). The gloss trap. In Z.-H. Han & T. Cadierno (Eds.), *Linguistic relativity in SLA: Thinking for speaking* (pp. 102-124). Clevedon: Multilingual Matters.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643-662. doi: 10.1037/h0054651

Taylor, L. J., & Zwaan, R. A. (2012). Fault-tolerant comprehension. In Y. Coello & A. Bartolo (Eds.), *Language and Action in Cognitive Neuroscience* (pp. 145-158). East Sussex: Psychology Press.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P.,…, & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*, 273-281. doi: 10.1162/0898929053124965

Theios, J., & Amrhein, P. C. (1989). Theoretical analysis of the cognitive processing of lexical and pictorial stimuli: reading, naming, and visual and conceptual comparisons. *Psychological review*, *96*, 5-24. doi: 10.1037/0033-295X.96.1.5

Tracy, R., & Gawlitzek-Maiwald, I. (2000). Bilingualismus in der frühen Kindheit. In H. Grimm (Ed.), *Enzyklopädie der Psychologie. Band 3: Sprachentwicklung* (pp.495-535). Göttingen: Hogrefe.

Van Elk, M., Slors, M., & Bekkering, H. (2010). Embodied language comprehension requires an enactivist paradigm of cognition. *Frontiers of Psychology*, *1.* doi: 10.3389/fpsyg.2010.00234

Verhelst, N., Van Avermaet, P., Takala, S., Figueras, N., & North, B. (2009). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience & Biobehavioral Reviews*, *35*, 407-426. doi: 10.1016/j.neubiorev.2010.04.007

Vukovic, N. (2013). When words get physical: evidence for proficiency-modulated somatotopic motor interference during second language comprehension. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3711-3716). Austin, TX: Cognitive Science Society.

Wang, X., & Forster, K. I. (2010). Masked translation priming with semantic categorization: Testing the Sense Model. *Bilingualism: Language and Cognition*, *13*, 327-340. doi: 10.1017/S1366728909990502

Willems, R. M., Toni, I., Hagoort, P., & Casasanto, D. (2010). Neural dissociations between action verb understanding and motor imagery. *Journal of Cognitive Neuroscience*, *22*, 2387-2400. doi: 10.1162/jocn.2009.21386

Willems, R. M., Labruna, L., D'Esposito, M., Ivry, R., & Casasanto, D. (2011). A functional role for the motor system in language understanding: Evidence from theta-burst transcranial magnetic stimulation. *Psychological Science, 22*, 849-854. doi: 10.1177/0956797611412387

Willems, R. M., Hagoort, P. & Casasanto, D. (2009). Body-specific representations of action verbs: Neural evidence from right- and left-handers. *Psychological Science, 21,* 67-74. doi: 10.1177/0956797609354072

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*, 625-636. doi: 10.3758/BF03196322

Yee, E., Chrysikou, E.G., Hoffman, E., & Thompson-Schill, S.L. (2013). Manual experience shapes object representations. *Psychological Science*, *24*, 909-919. doi: 10.1177/0956797612464658

Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. In D. Pecher & R. A. Zwaan (Eds.) *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 224-245). Cambridge: Cambridge University Press.

Zwaan, R.A., Stanfield, R.A., & Yaxley, R.H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*, 168-171. doi: 10.1111/1467-9280.00430

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *Journal of Experimental Psychology: General, 135*, 1-11. doi: 10.1037/0096-3445.135.1.1

Zwaan, R.A., & Yaxley, R.H. (2003). Spatial iconicity affects semantic relatedness judgements. *Psychonomic Bulletin & Review*, *10*, 954-958. doi: 10.3758/BF0319655

Supplementary Material

Language Categorization

Language group categorization		
Similar OL	Dissimilar-OL	Not classifiable
Albanian	Bahasa	Egyptian-Arabic
Bosnian	Japanese	Libyan-Arabic
Chechen	Korean	Marrokanisch-Arabisch
Croatian	Lingala	Eritrean
Dari	Pashtu	Ghanaian
English	Philippine	Kotokoli
Farsi	Swahili	Nigerian
French	Thai	
Greek	Turkish	
Hungarian	Urdu	
Italian		
Kurdish		
Modern Standard Arabic		
Polish		
Portuguese		
Punjabi		
Romanes		
Rumanian		
Russian		
Serbian		
Slovakian		
Syrian-Arabic		
Tunisian-Arabic		
Twi		
Vietnamese		