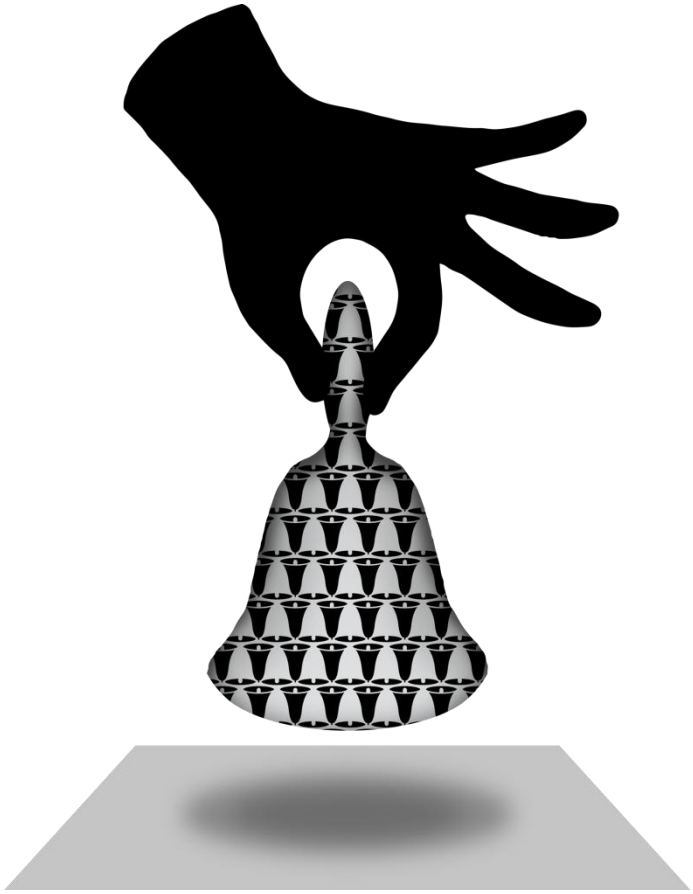
Ph.D. thesis

Understanding the world through L1 and L2

German and Spanish - The case of placement events



Dietha Koster
Department of Language and Communication
Faculty of Humanities
University of Southern Denmark

September 2016

I thank Vladimir Stankovic for designing the cover illustration. The illustration cleverly depicts the two main topics of this Ph.D. thesis. First, it shows a bell being placed onto a surface by a human agent, which symbolizes the central event type under investigation in this thesis that is a placement event. Second, within the large bell that is being placed onto the surface, you can distinguish small black and white bells depending on your visual focus. This depicts the notion that what you perceive depends on your perspective. In this thesis I have investigated whether our perspective may depend on the language we speak.

Acknowledgements

Three years ago I started as an "Early Stage Researcher" within the Marie Curie Initial Training Network on Language and Perception. I had a master's degree in Applied Linguistics, a curious mind and was enthusiastic to start a new adventure, yet I had never heard about something as "cognitive science" and knew little about "experimental designs" and "analyses of variance". Thus, there was a long way to go and I was in for a bumpy ride. This thesis reflects the distance I have covered, making small, sometimes larger steps toward personal and professional development. I realize no groundbreaking scientific advances have been made on the way. However, I hope that the ideas, data and arguments presented in this thesis will at least make a small contribution to our knowledge of the mono- and bilingual mind.

The realization of this project would have been impossible without the support of inspiring people around me, only some of whom I can mention here.

First of all I thank my main supervisor Teresa Cadierno. Her smart, lively personality and broad knowledge have benefited me in many ways throughout the project. Moreover, her detailed comments and deletion of redundant commas have greatly improved my writings. Second, I thank my co-supervisor Kenny Coventry, whose optimism worked uplifting. I have occasionally left meetings with Kenny with a migraine, but I learned much about psychology, experimental design and statistics. Third, I must thank Marco Chiarandini, who is not only a beautiful salsa dancer, but a multilevel regression hero in fact.

I thank my colleagues and friends in Denmark at the Department of Language and Communication. In particular, I thank Peiwen Li and Moiken Jessen, without whom I would have not survived the first year; my sweet "roomies" Marit Clausen and Annette Esbensen (who brilliantly helped to translate the Danish thesis summary) and Danica Jovanovic and Sima Erami, who continue to bring sunshine and laughter. A special word of thanks goes to Uwe Kjær Nilsen and Klaus Geyer for incessant kindness and help with Spanish and German translations.

Also, I thank my international colleagues. I thank the professors and fellows within the Marie Curie Network, from whom I have learned a great deal and whose company made for pleasurable attendance at the many Marie Curie events. I am also especially indebted to the professors, PhD colleagues and student-assistants I spent time with during my stays at University of East Anglia and Seville University; and outside the Marie Curie network at Bremen University, Humboldt University Berlin, University of Granada and University of Münster. In particular, I thank Harmen Gudde for many digital discussions, patient explanations of and smart insights into statistics.

Without the many partaking monolingual and bilingual students this project would have not been completed. I often personally plucked them from university squares or out of canteens and classrooms speaking only basic Spanish or German with a Dutch accent (though I improved over time), but I also received some great help from: Christian Freksa (University of Bremen); Myriam Pradillo Guijarro and Mercedes Fornés Guardia (*Instituto Cervantes*); Regine Höschen (*Goethe Institut*); Mariana Alvarez (*Sprachenzentrum*, Humboldt University Berlin); Stefan Bruns (VHS Mitte, Berlin); Alejandro Castañeda Castro (University of Granada); David Saldaña, Isabel Rodríguez Ortiz and Christoph Ehlers (University of Seville); Jens Bölte and Sarah Schimke (University of Münster) and Olga Gaudioso Solsona (*Sprachenzentrum*, University of Münster). Thanks to all for your effort and interest in this study.

I thank my beautiful friends with whom I enjoyed good times and relaxation next to PhD life. I thank both new friends I met in the last three years, Davide, Christian, Anita, Nathalie and Stephan, as well as friends who have accompanied me for quite a while now, Yfke, Rinske, Annemieke, Marij, Theresa, Julia and Hiske, and who will accompany me for much longer, I hope. Thank you for your support over Skype and in person, with visits to Denmark or whatever European city I was situated. I am also grateful for the wonderful salsa and lindy hop dancers with whom I have danced all over Europe in the past three years.

To conclude, I thank my sweet parents, Harry and Sieni, my dear sister, Margaretha and Manuel. Harry, who carried a mattress through the Berlin subway system to improve sleeping conditions in my temporary home. Sieni, who, despite fear of travelling, got on a train and stayed with me in Denmark through a tough period of time. Margaretha, with whom I can Skype for hours and who makes me smile with her original ideas. Manuel, who keeps me grounded, lifts me up and who is the most beautiful Spanish-German bilingual I know. Thank you for being you and for being there. Your incessant support means the world to me.

Table of Contents

English summary

This thesis investigates how individuals understand so-called placement events through their native (L1) or second (L2) language. Placement events are events where an agent moves an object to a certain location, as in: *He puts the book on the shelf*. The key motivation to study this topic is as follows. Actions of "putting" and "taking" are an ubiquitous part of everyday human experience and "putting" and "taking" verbs are among the most frequent and earliest learned verbs in a language. However, speakers of different languages employ different kind of verbs to describe placement events (Kopecka & Narasimhan, 2012). Drawing on the Sapir-Whorf hypothesis (Whorf, 1956), this leads one to wonder whether such cross linguistic differences affect how L1 speakers of different languages understand placement events. In extension, an interesting question is whether changes in an individual´s language system - by learning an L2 – can affect how s/he understands placement events.

Approximately 360 participants participated in the studies in this thesis: 60 L1 speakers of German, 60 L1 speakers of Spanish, 120 German learners of L2 Spanish and 120 Spanish learners of L2 German. The L2 learners were learning their L2 in a foreign language context (e.g. German learners of L2 Spanish in Germany). These adult L2 learners acquired the L2 after the age of 12 or post-puberty.

The aim of this thesis was to advance theories on language and perception. Therefore, we studied cross linguistic differences in the expression of placement events from three major theoretical perspectives on language and perception. Results are reported in three separate research papers. First, we studied placement events in light of the Sapir-Whorf (Whorf, 1956) and the Thinking-for-Speaking (Slobin, 1996) hypotheses. In particular, we investigated whether cross linguistic differences affect how individuals categorize (Paper 1) and memorize (Paper 2) aspects within placement events. We presented them with pictures depicting placement events (Paper 1) or sentences describing placement events followed by depicted placement events (Paper 2) and investigated whether language affected their perception of object orientation and gender of agents. In Paper 1, we investigated L1 German and L1 Spanish speakers. In Paper 2, we compared Spanish learners of L2 German and German learners of L2 Spanish with L1 German and Spanish control speakers.

The results of these studies show the following. In Paper 1, we found that in a context with no overt language use, cross linguistic differences did not affect how speakers categorize object orientation or gender of agents. Moreover, this study showed that although cross linguistic differences between languages exist, speakers may employ alternative linguistic strategies that result in similar descriptions of object orientation and gender of agents across languages. In Paper 2, we found that language affected perception of object orientation. We found that L1 German speakers had better recognition memory for object orientation than L1 Spanish speakers. When Spaniards learned L2 German and performed the task in German, their recognition memory for object orientation improved and was as good as that of L1 German speakers. When Germans learned L2 Spanish and performed the task in

Spanish, their recognition memory of object orientation was similar to L1 Spanish speakers´ memory. We found no effects for gender of agents.

Second, we examined placement events from a grounded cognition perspective (Paper 3). In particular, we investigated whether L1 and L2 speakers make "mental simulations" during language comprehension (Barsalou, 1999). We presented them with sentences describing placement events, which contained language-specific forms (verbs, suffixes) and investigated whether this led them to simulate object orientation and size as shown by so-called "match effects". In Paper 3, we compared Spanish learners of L2 German and German learners of L2 Spanish with L1 German and Spanish control speakers.

The results of this study show the following. We found no support that L2 readers simulate object orientation through German placement verbs. However, we did find support that L2 readers of Spanish augmentative suffixes make simulations of object size. In addition, we found that L2 readers process meaning slower than L1 readers.

Third, we investigated whether L2 learner differences affected (our measures of) their behavior in the studies reported in Paper 2 and 3. In particular, we investigated the role of L2 proficiency and the related factors L2 exposure and motivation. In Paper 2 we found that differences in general L2 proficiency did not cause differences in L2 learners´ memory. In Paper 3, we found that differences in specific L2 proficiency affected how fast L2 learners processed meaning. In addition, we found that the amount of L2 exposure affected how fast L2 German learners processed meaning.

In its entirety, this thesis contributes to theoretical advance in the following ways. We have expanded on three theoretical perspectives on language and perception, investigating a single domain of investigation, which is placement. First of all, our research provided evidence against the Sapir-Whorf hypothesis, but in favor of TFS effects on L1 and L2 speakers´ memory. Moreover, we found that L2 learners´ memory may differ from that of L1 speakers, irrespective of L2 proficiency. Thus, TFS effects in L1 and L2 speakers' memory seems an interesting topic for further investigation. Second, we found partial evidence in favor of theory on mental simulation in L1 speakers and L2 learners. On the one hand, this questions whether humans routinely make mental simulations for all types of object properties. On the other hand, there are several accounts that explain why simulation effects may be found or not. Thus, further research is needed to determine if and when simulation occurs. Third, we found that general L2 proficiency did not affect memory performance in Paper 2. Specific measures of L2 proficiency did reveal RT differences in L2 learner´ behavior in Paper 3. Amount of L2 exposure only reliably affected L2 German learners' speed of reaction in Paper 3, but not L2 Spanish learners' speed of reaction. There are two ways to interpret these results. First, it may be that in relatively simple language tasks, such as reading sentences, general language background differences do not affect L2 learners' behavior, thus no effects exist. Second, it may be that (our) measures of general L2

proficiency, L2 exposure and motivation are not fine-grained enough to convincingly reveal differences in L2 sentence comprehension.

All in all, we have shown that despite the universality of actions of placement, cross linguistic differences may indeed affect how L1 speakers understand placement events *if critical language is present*. Moreover, how L2 speakers understand placement events may change following their L2. General L2 learner differences do not convincingly seem to affect speed of meaning processing or memory accuracy. Only specific measures of L2 proficiency revealed differences in meaning processing. Therefore, the following lines of research seem more promising in making theoretical advance. First, more research comparing L1 and L2 speakers' TFS and its effect on memory is needed. Second, mental simulation theory needs to be further evaluated. In this quest, different conceptual domains (gender and size) and novel experimental tasks (such as the memory task employed in Paper 2) should be considered to determine whether and how effects of language on perception structurally occur.

Dansk resumé

Denne afhandling undersøger hvordan individer forstår *placement events* gennem deres førstesprog (L1) eller andetsprog (L2). *Placement events* er handlinger, hvor en agent flytter et objekt til et givent sted, som i: ”*Han sætter bogen på reolen*”. Motivationen for at undersøge dette emneområde er som følgende. Handlinger som ”at sætte” eller ”at tage” er en almindelig del af den menneskelige hverdag. Verberne ”sætte” og ”tage” er blandt de mest frekvente og tidligst tilegnede verber i et sprog. Det interessante er dog, at sprogbrugere med forskellige sprog anvender forskellige typer af verber for at beskrive *placement events* (Kopecka & Narasimhan, 2012). Med udgangspunkt i Sapir-Whorf hypotesen (Whorf, 1956) kan man undre sig over, om sådanne lingvistiske forskelle mellem sprogene influerer på, hvordan forskellige L1 sprogbrugere forstår *placement events*. I forlængelse heraf er det et interessant spørgsmål om forandringer i et individs sprogsystem – ved at tilegne sig et andetsprog (L2) – kan påvirke hvordan han/hun forstår *placement events.*

I afhandlingens undersøgelser deltog omkring 360 deltagere: 60 L1 tysktalende sprogbrugere, 60 L1 spansktalende sprogbrugere, 120 L1 tysktalende sprogbrugere med spansk som andetsprog (L2) samt 120 spansktalende sprogbrugere med tysk som andetsprog (L2). Andetsprogsbrugerne tilegnede deres L2 i en for dem fremmed sproglig kontekst (f.eks. tysktalende sprogbrugere med spansk som andetsprog (L2) i Tyskland). Disse voksne andetsprogsbrugere tilegnede deres L2 efter 12-årsalderen eller efter pubertetsalderen.

Formålet med denne afhandling var at videreudvikle teorier om sprog og perception. Med udgangspunkt i tre omfattende teoretiske perspektiver på sprog og perception undersøgte vi derfor lingvistiske forskelle på tværs af sprog i forhold til hvordan *placement events* udtrykkes. Resultaterne er rapporteret i tre separate forskningsartikler. For det første undersøgte vi *placement events* med udgangspunkt i hypoteser fremsat af Sapir-Whorf (Whorf, 1956) og Thinking-for-Speaking (Slobin, 1996). Mere detaljeret undersøgte vi, om forskelle mellem sprog påvirker hvordan individer kategoriserer (artikel 1) og husker (artikler 2) aspekter fra *placement events.* Vi præsenterede dem for billeder, der illustrerede *placement events* (artikel 1) eller sætninger, der beskrev *placement events* efterfulgt af illustreringer af *placement events* (artikel 2) og undersøgte om dette påvirkede deres perception i forhold til objektorientering og agenternes *gender*. I artikel 1 undersøgte vi tysk og spansktalende sprogbrugere (L1). I artikel 2 sammenholdte vi spansktalende sprogbrugere med tysk som andetsprog (L2) og tysktalende sprogbrugere med spansk som andetsprog (L2) med tysk og spansktalende (L1) kontrolpersoner.

Undersøgelsens resultater viser følgende. I artikel 1 fandt vi, at i en kontekst uden produktiv eller receptiv sprogbrug påvirkede forskellene mellem sprogene ikke måden, hvorpå individerne kategoriserede i forhold til objektorientering eller agenternes *gender*. Endvidere viste denne undersøgelse, at selvom der er forskelle mellem sprogene, så kan individerne på tværs af sprogene anvende alternative sproglige strategier resulterende i lignende beskrivelser af objektorientering eller

agenternes *gender*. I artikel 2 fandt vi, at sprog påvirkede perceptionen af objektorienteringen. Vi fandt, at L1 tysktalende sprogbrugere klarede sig bedre end L1 spansktalende sprogbrugere i forhold til at genkende objektets orientering. Når de spansktalende sprogbrugere med tysk som andetsprog (L2) udførte undersøgelsen på tysk, blev deres evne til at genkende objektets orientering forbedret og var ligeså god som de tysktalende sprogbrugeres (L1) evne til at genkende objektets orientering. Når tysktalende sprogbrugere med spansk som andetsprog (L2) udførte undersøgelsen på spansk, var deres evne til at genkende objektets orientering sammenlignelig med de spansktalende sprogbrugeres evne til genkendelse. Vi fandt ingen indvirkning i forhold til agenternes *gender*.

For det andet undersøgte vi *placement events* med udgangspunkt i et *grounded cognition* perspektiv (artikel 3). Mere præcist undersøgte vi, om L1 og L2 sprogbrugere danner *mental simulations* under sprogforståelsen (Barsalou, 1999). Vi præsenterede dem for sætninger som beskrev *placement events.* Sætningerne indeholdte sprogspecifikke elementer (verber, suffiks) og undersøgte om det fik dem til at simulere objektorientering og -størrelse i overensstemmelse med såkaldte *match effects*. I artikel 3 sammenholdte vi spansktalende sprogbrugere med tysk som andetsprog (L2) og tysktalende sprogbrugere med spansk som andetsprog (L2) med tysk- og spansktalende kontrolpersoner.

Resultaterne fra denne undersøgelse viser følgende. Vi fandt intet belæg for, at L2 sprogbrugere simulerer objektorienteringen gennem tyske *placement* verber. Vi fandt dog, at når L2 sprogbrugere læser sætninger med ”*augmentative suffixes”*, suffiks der indikerer stor størrelse, simulerer de objektstørrelse. Yderligere fandt vi, at L2 sprogbrugere processerer *meaning* med en langsommere hastighed i forhold til L1 sprogbrugere, når de læser de præsenterede sætninger.

For det tredje undersøgte vi, om forskellene mellem andetsprogsbrugerne (L2) påvirkede (vores mål anvendt til bedømmelse af) deres adfærd ved deltagelse i undersøgelserne rapporteret i artikel 2 og 3. Mere præcist undersøgte vi L2 færdigheder, L2 eksponering og motivation til at lære L2. I artikel 2 fandt vi, at forskelle i generelle L2 færdigheder ikke medførte forskelle i forhold til L2 sprogbrugernes hukommelse. I artikel 3 fandt vi, at forskelle i specifikke L2 færdigheder påvirkede hvordan L2 sprogbrugerne processerede *meaning*.

I sin helhed bidrager denne afhandling til teoretisk videreudvikling på følgende måder. Vi har udvidet vores viden i forhold til tre teoretiske perspektiver for sprog og perception ved at undersøge et enkelt domæne, nemlig *placement*. For det første bidrager vores forskning med resultater, der ikke bekræfter Sapir-Whorf hypotesen, men derimod er i overensstemmelse med, at TFS påvirker L1 og L2 sprogbrugernes hukommelse. Yderligere fandt vi, uafhængig af L2 færdighederne, at L2 sprogbrugernes hukommelse er forskellig fra L1 sprogbrugernes. Derfor lader TFS’s indvirkning på L1 og L2 sprogbrugernes hukommelse til at være et interessant emne for videre fremtidig forskning. For det andet fandt vi delvis belæg for teorien om *mental simulation* i forhold til L1 og L2 sprogbrugere. På den ene side stiller det spørgsmålstegn ved om individer rutinemæssigt laver *mental simulations* for alle typer af objektkarakteristika (f.eks. orientering, størrelse og farve). På den anden side er der andre

studier, der indikerer ingen *simulation effects*. Derfor er det nødvendigt med en nærmere undersøgelse for at konkludere om og hvornår *simulation* forekommer. For det tredje fandt vi, at generelle L2 færdigheder ikke påvirkede præstationen i forhold til hukommelse i artikel 2. Specifikke undersøgelser af L2 færdigheder afslørede dog en forskel i reaktionstiden i forhold til L2 sprogbrugernes adfærd i artikel 3. Graden af L2 eksponering påvirkede kun reaktionstiden hos L2 tysktalende sprogbrugere i artikel 3, men ikke hos L2 spansktalende sprogbrugere. Der er to måder at fortolke disse resultater på. For det første kan det være, at forskelle i den sproglige baggrund ikke påvirker L2 sprogbrugernes adfærd i relativt simple sproglige opgaver som f.eks. det at læse sætninger. For det andet, kan det være at (vores) mål for undersøgelse af generelle L2 færdigheder, L2 eksponering og motivation ikke var detaljeret nok i forhold til, at kunne vise forskelle i L2 sprogbrugernes sætningsforståelse.

Alt i alt har vi vist at forskelle mellem sprog, uafhængigt af de universelle handlinger af *placement*, i særdeleshed påvirker måden hvorpå L1 sprogbrugerne forstår *placement events*. Det kommer kun til udtryk, når sprogbrug med verber og suffiks bliver anvendt. Endvidere kan sprogforståelsen ændre sig for L2 sprogbrugere i kraft af deres L2. Generelle forskelle mellem L2 sprogbrugerne lader ikke til at påvirke processeringshastigheden i forhold til *meaning* eller hukommelsespræcision. Kun specifikke undersøgelser af L2 færdigheder afslørede forskelle i forhold til processering af *meaning*. I forhold til fremtidig teoretisk videreudvikling er der brug for en undersøgelse af følgende. For det første er der brug for mere forskning i forhold til sammenligninger af L1 og L2 sprogbrugernes TFS og dens indvirkning på hukommelse. For det andet er der brug for en nærmere evaluering af teorien bag *mental simulation.* I forlængelse heraf bør det overvejes, hvordan forskellige konceptuelle domæner (f. eks. *gender* og størrelse) samt nye eksperimentelle opgaver (som f.eks. hukommelsesopgaverne i artikel 2) anvendes i forhold til at konkludere om og hvordan sprog rent strukturalistisk påvirker perception.

Chapter 1: Introduction

The aim of this chapter is to explain how and in which context this thesis came into being. I provide information on the larger project of which this subproject formed part. Also, I discuss the motivation for and focus of this project and present the collaborators on the papers presented in this thesis. Finally, I present the outline of the thesis and clarify key terms that will be used throughout the dissertation.

1.1 Background: Language and Perception

The research reported in this dissertation has been supported by the EU 7th Framework Programme Marie Curie Initial Training Networks grant Nr. 316748 under the project *Language and Perception*.
I was involved in this network as an "Early Stage Researcher" (ESR). The goal of the network was to provide partakers a unique approach to understanding the interaction between two central cognitive systems: language and perception. Traditionally, researchers have studied these systems independently, and received training from a single discipline perspective. The Marie Curie Network, for the first time, offered researchers training in an interdisciplinary approach to the examination of the bidirectional relationships between language and perception. The Network supported both basic and applied research in both clinical settings and in industry. The work packages in the training program formed three interconnected clusters:

1. *Language-perception interactions in healthy participants - basic research* investigates theories of language and perception from complementary perspectives in order to break new ground in understanding language-perception interactions.
2. *Language-perception interactions in atypical populations – applied research* transports state-of-the-art methods of language-perception to atypical populations in order to identify underlying mechanisms for different kinds of atypical behaviors.
3. *Advanced technologies for language and perception research* bridges basic research and its applications in the development of new behavioral and neurophysiological techniques to identify the interaction between language and perception among typical (adults and children) and atypical populations.

The Network started in September 2013, with an opening workshop at the University of Aston in Birmingham and was concluded with the Language and Perception conference at NTNU in Trondheim, June 2016. In the three years in between, researchers and industrial partners met on nine trainings or conferences organized by the different European institutes involved. Moreover, the network provided opportunities for researchers and industrial partners to collaborate outside these official events. Examples of such international collaboration were the "secondments" that the ESRs did. During these secondments, ESRs stayed for an extended period at one or more partner institutions to exchange knowledge and resources and experience working in different institutes. ESRs were encouraged to

conclude their training period with a dissertation aimed to obtain the Ph.D degree, to be submitted at the universities where they had been placed.

1.2 The current Ph.D Project

The current project was carried out within the framework of the first cluster identified in the Marie Curie training program. Thus, the general goal was to investigate psycholinguistic theories of language and perception in order to break new ground in understanding language-perception interactions in healthy adults. The specific goal was to investigate these interactions in monolingual (L1) speakers versus second language (L2) learners in a foreign language context (e.g. Spanish learners of German in Spain). These L2 learners are typically classroom learners, who learn the L2 via the means of the first (Weinreich, 1953). Studying L2 learners, or bilinguals, is relevant for at least two reasons. First, the bilingual mind constitutes a favorable ground for testing psycholinguistic theory, as bilinguals combine so to say "two languages in one mind" (Pavlenko, 2014). Thus, studying language-perception interaction in bilinguals may provide for new insights and theoretical advances in cognitive science. Second, in a time of globalization, transnational migration and increased ethnolinguistic diversity in the world, it is critical to understand how the human mind is affected by multilingualism. In this way, knowledge-based decisions can be made considering language policy, language education and the development of L2 teaching materials.

The focus of investigation was a specific type of motion event, a so-called "placement event". A placement event is an event where an agent moves an object to a certain location, as in: *He puts the book on the shelf*. This may seem a rather narrow and specialized area of interest, but there are several reasons for this choice (Levinson, 2012). One of the most important reasons is that simple actions of putting and taking things from places are a ubiquitous part of everyday human experience. Thus, it is not surprising that verbs of "putting" and "taking" are amongst the most frequent, basic verbs in a language and that they are amongst the earliest verbs learned by children. However, if we want to explain how we conceptualize putting and taking actions for purposes of language, we encounter profound intellectual challenges (Narasimhan et al., 2012). We find that different world languages display a curious amount of variation in the use of positional verbs to describe putting and taking actions, varying from a tight closed obligatory set of 3-5 verbs, to a much wider set of 12-20 or more verbs (see Ameka & Levinson, 2007). This variation offers an interesting laboratory to study the interaction between the basic cognition of reaching and placing and the corresponding linguistic description of such actions.

In this project we focused on the German and Spanish language, since these languages differ in how they express various aspects of placement scenes. Crucially, they show differences in the semantics of the verbs they employ to describe the action of putting (Lemmens, 2006; Cadierno et al., 2016). In addition, transcending yet relevant to placement events, German and Spanish also differ in the way

they mark gender of human agents (Cartagena & Gauger, 1989) and size of objects (Gooch, 1967; Lohde, 2006). We examined the implications of this cross linguistic variation for speakers' non-linguistic cognition in L1 speakers and adult L2 learners. Non-linguistic cognition involves cognitive processes such as mental imagery, categorizing and memorizing (Gardner, 1985).

In summary, cross linguistic variation in the expression of placement events in L1 and L2 German and Spanish and its effect on non-linguistic cognition stands central in each of the papers presented in this dissertation. However, each paper investigates a series of specific research questions. In Paper 1 the central question is whether cross linguistic variation affects how L1 German and Spanish speakers categorize position of objects and gender of agents in placement events. We also examine their linguistic descriptions of these events. In Paper 2 we examine whether L2 speakers´ (and L1 controls´) recognition memory is affected by aspects encoded in the L2 versus the L1. In addition, we ask whether potential effects are mediated by L2 proficiency. In Paper 3, we ask whether L2 speakers (and L1 controls) make mental simulations of object orientation and size. Also we examine whether L2 proficiency, L2 exposure and motivation to learn the L2 affect if and how L2 speakers process meaning.

1.3 Collaborators

This Ph.D. project has involved close collaboration with researchers with expertise in linguistics, psychology and statistics in terms of study design and data analysis. In addition, I received assistance from student-assistants and technicians with data collection. I have also received support from *Instituto Cervantes Bremen* and *Berlin* and *Goethe Institut Granada*; and university departments, language centers and language teachers in Bremen, Berlin, Münster, Granada and Seville, to find L1 and L2 participants. It follows that the articles presented in this dissertation are the culmination of these collaborations. In the following, the contributions of the co-authors of each article in this dissertation are briefly recounted.

Teresa Cadierno and Kenny Coventry, as the Ph.D. candidate´s main supervisor and co-supervisor respectively, are co-authors of Paper 2 and 3. They were involved in the first conceptualization of the experiments presented these papers. Throughout the design and conduction of the different studies they offered extensive advice and guidance. In addition, Kenny had a critical role in inspecting - and in some cases, improving - statistical analyses performed by the main author. In Paper 2 and 3 the main author wrote all sections of the article. Teresa and Kenny provided comments and suggestions that greatly influenced the shape and content of the papers.

Marco Chiarandini, as interdepartmental collaborator at the University of Southern Denmark, is co-author of Paper 3. He played a critical role in selecting the appropriate statistical tests in discussion with the main author. In addition, he performed the selected statistical analyses in the statistical

program R and discussed results in close collaboration with the main author. Finally, he provided comments and suggestions on several drafts made by the main author, which greatly influenced the results and conclusion section of Paper 3.

1.4 Outline Dissertation

This dissertation is organized as follows. In the first part of Chapter 2 I present the central theories that constituted the starting point of the particular studies presented in the separate articles. Here, I also discuss empirical studies into the particular linguistic and conceptual domains that were investigated in the different papers. In the second part of Chapter 2 I discuss in detail the cross linguistic differences under investigation in this thesis. Chapter 3 focuses on the methodological aspects of the three different studies, describing the samples of L1 speakers and L2 learners; experimental tasks and analytic techniques used in the different papers. In Chapter 4, I provide a summary of the findings in each paper and I review findings for each of the cross linguistic differences under investigation. In the concluding chapter, Chapter 5, I discuss the contributions and limitations of the papers in relation to the theories they sought to test or expand and I outline directions for future research.

Throughout the dissertation, the terms "monolingual" and "L1 speaker" are used to refer to speakers that almost exclusively use their native language on a day-to-day basis. Alferink and Gullberg (2013) refer to such speakers as "functional monolinguals".

The terms "bilingual" and "L2 learner" are used interchangeably. As many different types of bilinguals may be distinguished, when of relevancy, I specify in the separate papers what types of bilinguals are being discussed. As mentioned, the L2 learners under investigation in this dissertation are situated in a foreign language context and are "compound bilinguals" (Weinreich, 1853). This means that they are classroom learners, who are learning the L2 via the means of the first.

I understand "cognition" as "the acquisition, storage, transformation and use of knowledge" (Matlin, 2005). I understand "perception" as the organization, identification, and interpretation of sensory information in order to represent and understand the environment (Schacter, 2011).

2. Theoretical Background

The aim of this chapter is to present previous theoretical and empirical work of relevance to the work presented in this thesis. This chapter consists of two parts. In the first part, I present three theoretical paradigms that are central in this thesis. Here, I summarize the main characteristics of the main theories, review empirical evidence and point out theoretical challenges that will be addressed in this thesis. I also formulate the main research questions that will be addressed in the separate papers.
In the second part, I outline the cross linguistic differences in the conceptual test domain chosen in this thesis, namely "placement events". In particular, I discuss the meanings encoded in German and Spanish verbs, personal pronouns and suffixes that may be used in the description of placement events. I also discuss how salient these forms may be for language users and how they may be taught to L2 learners.

PART I:

This section presents an overview of three theoretical paradigms, which discuss language and thought in mono- and bilingual individuals. The goal is to concisely discuss the theoretic foundations and assumptions on which the current empirical research is based. Where appropriate, I relate the individual papers to the theory.

All papers in this thesis relate to the general theoretical challenge put by cognitive science. That challenge is "to characterize the nature of human knowledge - its forms and content - and how that knowledge is processed, acquired, used and developed" (Gardner, 1985). Cognitive science is a highly interdisciplinary field that emerged in the 1950s, combining theories and methods from psychology, computer science, linguistics, philosophy, neuroscience and anthropology (Gardner, 1985). In this dissertation we focus on human language as a cognitive system (Sloan Foundation Report "Cognitive Science", 1978).

Human linguistic communication differs from the communication of other animal species in at least three fundamental ways (Tomasello, 2003). First, human linguistic communication is symbolic – linguistic symbols are social conventions by means of which an individual attempts to share attention with another individual. Other animals do not communicate with one another using linguistic symbols. Second, human linguistic communication is grammatical. Human beings use their linguistic symbols in patterned ways, and these patterns, or linguistic constructions, take on meaning of their own. Third, unlike other animal species, the human species does not have a single system of communication. Different groups of human beings have conventionalized more than 6000 different systems of communication; and different human beings may acquire more than one system of communication (Tomasello, 2003).

There are different roads towards the conventionalization of a particular communication system (Kroll & De Groot, 2005). Children typically acquire only the system(s), their L1(s), of their natal group(s) through years of daily interaction with mature language users. However, mature language users may consciously choose to acquire other communication systems, a second (L2), third (L3) or even fourth (L4) language, even if they find themselves in an environment where the target language is not spoken (e.g. in a foreign language context). These differences in L1 and L2 language acquisition raise important theoretical questions. One important question is if and how the L1 and L2 affect other cognitive systems, such as perception (Miller & Johnson-Laird, 1976; Cook & Bassetti, 2011). Another central question is if and how mono- and bilingual language comprehension differ (Perfetti, 1999; Dijkstra, 2005). Finally, we may ask if individual differences in L2 speakers will affect perception and comprehension. Therefore, these three themes warrant discussion in relation to the current dissertation.

The following paragraphs describe three theoretical paradigms that are relevant to the empirical work presented in this dissertation. First, I discuss the Sapir-Whorf hypothesis and the Thinking-for-speaking (TFS) hypothesis (Whorf, 1956; Slobin, 1996). The main idea of both hypotheses is that language may affect other cognitive systems, such as perception, but the hypotheses differ in the degree of influence they attribute to language. In Paper 1 and 2 of this dissertation, we discuss The Whorfian and TFS hypotheses in relation to the cognitive processes of categorizing and memorizing. Second, I discuss symbolic and grounded approaches to language comprehension, with the latter stating that language is grounded in perceptual experience and action (Barsalou, 1999). As a consequence grounded approaches state that language comprehension is taken to involve "mental simulation" or a reactivation of traces of earlier interactions with the world (Barsalou, 2008; Glenberg, 1997; Glenberg & Robertson, 1999, 2000; Lakoff, 1987; Pecher & Zwaan, 2005). In Paper 3, this notion is examined in relation to L1 and L2 comprehension. Third, I discuss how L2 learner differences may affect cognition (Jarvis & Pavlenko, 2007). In particular, I discuss the factors L2 proficiency, L2 exposure and motivation. Also, I describe empirical studies that have started to explore effects of these or related factors on language processing, memory and language comprehension. In paper 2 and 3, we investigate such effects for our L2 learners.

2.1 Sapir-Whorf Hypothesis and Thinking-for-speaking

The idea that language may affect how individuals perceive the world is usually traced back to Benjamin Whorf (1956). The works of the latter, together with those of Edward Sapir (1929), led to the formulation of the Whorfian, or alternatively Sapir-Whorf hypothesis. This hypothesis states that (1) languages vary in their semantic partitioning of the world; (2) the structure of one´s language influences the manner in which one perceives the world; (3) therefore, speakers of different languages will perceive the world differently (Hoijer, 1954). In the 1950s and 1960s, the Whorfian position was embraced and supported by Brown and Lenneberg´s (1954) studies on color terms. However, negative findings in the same domain by Rosch (1973) introduced a period of skepticism about linguistic

influence on thought. The implication of Rosch´s findings was that perception of color is determined by the biology of human color perception and not by the language learned. The skepticism was reinforced by dominant ideas in adjacent fields, such as Chomsky´s (1980) notions of a universal grammar, separation of language from cognition and de-emphasis of semantics. In cognitive psychology and development, the dominant position was that human conceptual structure is relatively constant across cultures and that conceptual and semantic structures are closely coupled. There was a strong sense that concepts come first and language merely names them (Gentner & Goldin-Meadow, 2003).

In the last decades the language-and-cognition question has again become an area of active investigation. An important reason for this revival is the following. In the 1970s, semantic analyses of Talmy (1975), Langacker (1976), Bowerman (1980) and others showed that important differences exist in how different languages carve up the world. This showed one cannot maintain that conceptual structure is universal, if semantic structure reflects conceptual structure. Either, one must adopt that semantic and conceptual structure are independent of one another, leaving the universal view intact. Alternatively, one may choose to explore whether semantic structure can influence conceptual structure. An important line of research chose the latter, moving away from the focus on color to the study of domains such as space. Spatial relations are highly variable cross linguistically (Bowerman, 1980; Casad & Langacker, 1985; Talmy, 1975, 1985), which suggests the possibility of corresponding cognitive variability. The work of Stephan Levinson´s group on cognitive differences that follow from differences in spatial language has been influential in attracting renewed interest to the Whorfian question (e.g. Levinson, 1996; Li & Gleitman, 2002).

The Whorfian hypothesis has been interpreted in weaker and stronger manners (see Wolff & Holmes, 2011 for a review). A weaker interpretation of the Whorfian stance is Slobin´s (1996) thinking-for-speaking (TFS) hypothesis. Slobin´s critical refinement was that language may affect thought when one is thinking with the intent to *use* language. In his 1996 paper, he mainly speaks about the study of "mental processes that occur during the act of formulating an utterance", thus speaking. The focus on speaking may result from his extensive study of the verbalization of motion events in the famous frog stories by speakers of different languages (Berman & Slobin, 1994). Yet in 2003, next to speaking, Slobin poses that he regards all forms of linguistic production (speaking, writing, signing) and reception (listening, reading, viewing); as well as a range of mental processes (understanding, imaging, remembering etc.) as TFS. Focusing on language and memory, he argues that "many of the events that we remember are encountered only through narrative […] It is quite likely that the language in which information is presented – both fictional and documentary – plays a role in the ways in which information is stored and evaluated" (Slobin, 2003: 177).

In the 1960 and 70s researchers discovered that bilinguals constitute a favorable ground for testing Whorfian or TFS hypotheses (Carroll, 1956, Macnamara, 1970, 1991). Early empirical work by Ervin(-Tripp) showed that bilinguals´ cognitive behavior shifted with a change of language (Ervin,

[1953]1973, Ervin-Tripp, 1967). Similar findings came from Lambert and colleagues (Jacobovits & Lambert, 1961, Lambert, Havelka & Crosby, 1958, Lambert & Rawlings, 1969). However, bilingualism also appeared as support for anti-Whorfian arguments (Macnamara, 1970, 1991; Gleitman & Papafragou, 2005). In the 1980s and 90s bilingualism was consolidated as a field of research, with publication of several foundational texts (Baker, 1993; Baker & Prys-Jones, 1998; Grosjean, 1982; Romaine, 1989). Therefore, in the past decade, bilinguals have become the focus of more systematic explorations of language and cognition. New professional forums facilitated discussion of applying theories and methods developed in neo-Whorfian inquiry to the study of the bilingual mind. This has resulted in several volumes on bilingualism and cognition (Cook & Bassetti, 2011, Han & Cadierno, 2010, Jarvis, 2011; Jarvis & Pavlenko, 2007, Pavlenko, 2011).

Currently, the literature is unresolved about if and in which contexts the Sapir-Whorf hypothesis holds (Trueswell & Papafragou, 2010; Fausey & Boroditsky, 2011; Bylund and Athanasopoulos, 2014). It has been suggested that whether language effects on cognition occur depends on a number of factors such as the specific characteristics of the domain involved, the nature of the particular linguistic feature under investigation and the degree to which an experimental task promotes or inhibits strategic use of linguistic categories (Bylund & Athanasopoulos, 2014).

2.1.1 Empirical Evidence for and against Sapir-Whorf and TFS hypotheses

There is considerable empirical work on Whorfian and TFS questions. Most of the work has been conducted with L1 speakers, but in the past years the number of studies that study bilinguals has been increased. A typical monolingual Whorfian or TFS setup involves speakers of two different world languages (for example, German versus Spanish). One language group (that linguistically expresses aspects of a certain domain) is expected to show a Whorfian or TFS effect whereas the other "control" group (speaking a language that does not linguistically express aspects of the same domain) is expected to show no effect. Below I give a limited overview of studies into the domains of (motion in) space and gender, which are relevant domains within this thesis. Size is also a relevant domain in this thesis, yet not in direct relation to Whorfian or TFS paradigms. Thus, in particular, I discuss space and gender studies that employed experimental tasks relevant to those used in the current thesis.

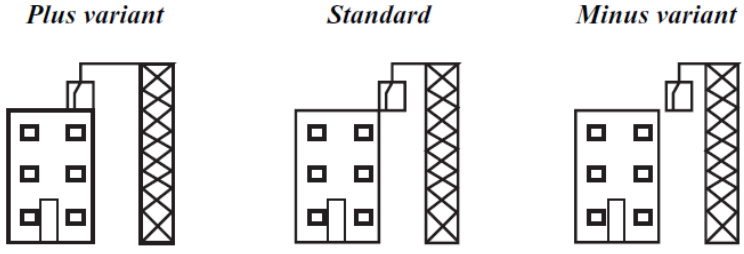
Within the Whorfian and TFS tradition, categorization tasks have been a popular means to test for language effects on cognition. Categorization is considered to operate on the basis of similarity, so that two stimuli that are perceived as similar are likely to be classified as members of the same category (Nosofsky, 1986). Therefore, categorization tasks measure "cognitive preferences" (Boroditsky, Ham & Ramscar, 2002). In categorization tasks participants are oft presented with a set of three pictures (as in Papafragou et al., 2002). Next, they are asked to choose two pictures that, according to them, are most closely related. In strictly Whorfian task versions there is no overt use of language while participants perform a task; in TFS task versions, participants receive or produce linguistic descriptions of pictures. If participants select the pictures that are linguistically related as most similar, it is taken as

support for the Whorfian or TFS view. A variation on this task is a task where two pictures´ similarity is rated on a Likert scale (as in Boroditsky et al., 2002). Here, high similarity ratings to linguistically related pictures and low ratings to linguistically unrelated pictures are interpreted as support for the Whorfian or TFS view.

In the domain of (motion in) space and gender, categorization studies show mixed results. Among other things, these studies differ with respect to the language-pairs they investigate; the timeframe in which stimuli are presented (simultaneous (all stimuli at the same time), subsequently (one stimuli after the other) and stimuli type (static or dynamic; black-and-white or color). Studies on motion verbs that encode manner of motion have yielded no Whorfian effects with L1 speakers (Gennari et al., 2002, Papafragou et al., 2002, Cardini, 2010 and Trueswell & Papafragou, 2010). However, where there is (the possibility of using) language in the task, Gennari et al. (2002) and Trueswell & Papafragou (2010) find effects of language on both categorization and memory. The possibility to use language comes in tasks with so-called "free encoding" conditions. Here, participants study stimuli without any instructions, which means they could employ language as a strategy to complete a task or not. Hae In Park and Ziegler (2014) report effects of "put in" and "put on" with a free encoding categorization task, both for L1 and L2 speakers. Athanasopoulos et al. (2015) also showed Whorfian effects for categorization of motion events for bilinguals with free encoding. In the case of gender, studies on gendered articles (presented with object nouns) yielded Whorfian effects in adult and children L2 speakers (Sera et al., 1994) and L1 and L2 speakers (Philipps & Boroditsky, 2003; Boroditsky, Schmidt & Philipps, 2003).

Another Whorfian or TFS line of research has investigated the effects of language on memory and perception. Feist and Gentner (2007) employed an interesting "picture recognition" task. In this task, participants are first presented with a drawing of a spatial scene (Figure 1, Standard), with or without a sentence describing the spatial relation through a spatial preposition (e.g. *the block is on the building*). Second, participants see a drawing that is either identical or pictures displaying a stronger (Figure 1, Plus variant) or weaker (Figure 1, Minus variant) version of the spatial relation seen in the original picture. Then participants are asked to indicate whether this picture was identical as compared with the preceding picture. The hypothesis was that participants would falsely recognize stronger versions as identical to the original picture when reading spatial prepositions.

Figure 1 (from Feist & Gentner, 2007): Examples of plus, standard and minus picture variants of spatial relationships and a sentence example containing a spatial preposition.



The block is on the building.

Coventry et al. (2010) conducted a study on English-Spanish cross linguistic variation in the expression of support and containment relationships, based on the study by Feist and Gentner (2007). They slightly changed Feist and Gentner´s task by including a control question at the first stage of the task. Participants were asked to judge whether a sentence with a fitted well with picture depicting stronger or weaker versions the spatial relationship. This control question was included to ensure that participants seriously studied sentences and pictures.

Regarding (motion in) space, the memory tasks described above or related tasks yield mixed results as well. Feist and Gentner found support in favor of TFS: when their L1 participants read spatial prepositions, they falsely recognized the *stronger* version as identical to the original picture. Coventry et al. (2010) intended to present their task to L2 learners. However, as they found that their L1 "control" speakers showed no language-perception effects, the L2 part of the study was not carried through. Other studies with related memory tasks focusing on space (motion) also yielded mixed results. For example, motion verbs did not affect memory in a Whorfian setup where no language was produced or encountered (Finkbeiner et al., 2002). Also, Bosse and Papafragou (2010) found no Whorfian or TFS effects testing the German verbs *sitzen* [to sit] and *liegen* [to lie]. However, Loftus and Palmer (1974) and Billman et al. (2000) found effects for motion verbs within a TFS paradigm. In addition, Filipovic (2011) reported TFS memory effects of motion verbs for bilinguals. She found that English-Spanish bilinguals´ descriptions of motion aligned with Spanish lexicalization patterns and L1 Spanish speakers´ recognition memory. This study suggests that L2 learners´ memory may be affected by the language they employ.

With respect to gender, we know of one empirical study that focused on memory. Boroditsky, Schmidt and Philipps (2003) showed that Spanish and German speakers´ memory for object-name pairs (e.g. *apple-Patricia*) was better for pairs where the gender of the proper name was consistent with the grammatical gender of the object name in their L1. The attribution of grammatical gender to the particular objects was reversed for Spanish and German. The effect was shown even though participants performed the memory task in English.

2.1.2 Theoretical Challenges

Empirical work has not allowed for conclusive thoughts on the validity of Whorfian or TFS effects. This holds for both the domain (motion in) space and gender. The question then is, How do we move forward? First, we note that there has been more work on space (motion) than on gender. Yet, the couple of studies on gender did yield support for Whorfian categorization and memory effects. We know of no study that has combined two domains (such as space and gender) within one investigation. Second, we note there has been more work focusing on monolinguals, than studies with bilingual speakers. Yet, the couple of studies with bilinguals yielded support for both Whorfian and TFS effects for memory. These observations may provide ways to make theoretical advance.

A first possible road to theoretical advance is to improve the design of categorization tasks. We observe that studies typically employ categorizations tasks that investigate one conceptual domain (e.g. space). We also observed there are two types of categorization tasks that may reveal fine-grained (similarity judgment on a scale) and coarser differences (forced choice) in categorization. Finally, we observed that the categorization of (motion in) space and gender in placement events has not been investigated yet (but see Bosse & Papafragou, 2010 for a related investigation on space indicated by German positional verbs). Taking into account these observations, first, we argue that strong support in favor or against Whorfian or TFS effects could be found if a study investigates multiple domains (e.g. space and gender) and language effects can be explained both ways (e.g. Language 1 shows an effect for space, but not for gender and Language 2 vice versa). Second, we reason that applying two different types of categorization tasks, similarity judgment by scale and categorization by forced choice, will provide us with both fine-grained and coarser categorization information.

In Paper 1 we take two domains, two categorization tasks and speakers of two different languages (e.g. German and Spanish) as a starting point. Language 1 (German) linguistically encodes object position whereas Language 2 (Spanish) does not. However, Language 2 linguistically encodes the gender of agents whereas Language 1 does not. In the experimental task, German and Spanish participants are presented with picture sets that can be related based either on German space or on Spanish gender. The hypothesis is that German speakers will give pictures that are spatially related high ratings; whereas Spanish speakers will rate pictures related on gender high; and vice versa. Also we hypothesize that German speakers will make forced categorization choices based on similarity in object position whereas Spanish speakers will make forced categorization choices based on similarity in gender of agents. Such a study design is able to provide double support in favor or against Whorfian effects. In Paper 1, we also collect linguistic descriptions of participants to examine whether our linguistic predictions are correct.

More work on L2 speakers may be the second option for theoretical advance. In particular, this holds with respect to the relation between language and memory, which has not been investigated abundantly in L2 speakers (as pointed out by Filipovic, forthcoming). If bilingual speakers convincingly show behavioral shifts with a change of language, this strongly suggests that language may indeed affect perception. This can be shown in a bidirectional study, where a comparison is made between L1 and L2 speakers of two languages (e.g. L1 speakers of German and Spanish; German learners of L2 Spanish and Spanish learners of L2 German). In Paper 2 of this thesis we set up a bidirectional TFS study with bilingual speakers. In this study we employ an experimental task as in Coventry et al. (2010).

2.1.3 Research Questions (Paper 1 and 2)

In sum, the main research questions that guide the research presented in Paper 1 and 2 are:

Paper 1:
How do German and Spanish speakers categorize placement events where object orientation and gender of agents are manipulated? Do cross linguistic differences in the domain of space and gender affect their categorization choices in a context without overt language use?

Paper 2:
How do L1 and L2 speakers of German and Spanish memorize object orientation and gender of agents in placement scenes? Do cross linguistic differences in the domain of space and gender affect their memory in a context with covert language use?

2.2 Symbolic and Grounded Cognition

Cognitive science has produced many theoretical accounts on how the human mind handles language. However, in the past decade there has been a shift in how cognition in general, and language in specific, is conceptualized. Up until fifteen years ago, the mainstream view was that the human mind manipulates abstract, amodal and arbitrary symbols (i.e. words) combined by syntactic rules (e.g., Burgess & Lund, 1997; Chomsky, 1980; Fodor, 2000; Kintsch, 1988, Pinker, 1994). This means that the human mind was assumed to work as how a computer works. The main problem with this conceptualization of cognition is that it has no connection to actual experience (the symbol grounding problem: cf. Harnad, 1990). Therefore, in recent years, theories under the term "grounded cognition" (Barsalou, 1999a, 1999b, 2008; Pecher & Zwaan, 2005 Glenberg, 1997; Glenberg & Robertson, 1999, 2000; Lakoff, 1987) or "embodied cognition" have proposed a different perspective. These theories propose that human thought and language are grounded in or shaped by our bodies and (perceptual) experiences with the world.

The idea of grounded cognition has inspired empirical work and yielded some hallmark findings. However, recently the validity of some hallmark findings is being challenged. An example of a hallmark finding is the Action-Sentence Compatibility Effect (ACE) (Glenberg & Kaschak, 2002). In the ACE, people process sentences implying movement toward or away from themselves, responding with actions toward or away from their bodies. Glenberg and Kaschak (2002) found that these processes interact, which implies a linkage between linguistic and motor systems. From a theoretical perspective this finding has been extremely influential as it has been widely cited in favor of embodied cognition. However, recently Papesh (2015) reported repeated failures to replicate the ACE. In addition, she performed a Bayes Factor analysis of previous ACE literature, which suggests that the evidence in favor of ACE is generally weak. In addition, Durgin et al. (2009) and Firestone and Scholl (2014; in press) have provided evidence against some other well-cited effects of action capabilities on perception. They suggest that previous studies have employed task demands that are not reflective of perception and thus do not allow an interpretation of results as top-down effects on perception.

Another hallmark notion in grounded cognition is that humans make "mental simulations" during language comprehension. The idea of mental simulation during language comprehension was put forward in a seminal paper by Lawrence Barsalou in 1999. Here he argued that perception (and possibly language comprehension) involves mental simulators that bind specific tokens in perception (i.e. individuals) to knowledge for general types of things in memory (i.e. concepts) (Barsalou, 1999: 581). Mental simulators perform this mental simulation by reactivating previous perceptual experiences with the world. Barsalou (1999: 586) noted that simulations may differ in terms of completeness. He points out that: ' (…) a simulator produces simulations that are *always* partial and sketchy, *never* complete. As selective attention extracts perceptual symbols from perception, it never extracts all the information that is potentially available.' This implicates that the nature of mental simulations may differ across situations and that in certain contexts, comprehenders may be motivated to construct more detailed simulations as compared with other contexts.

Zwaan (2014) argues that the contributions of abstract symbols and mental simulations to language comprehension vary as a function of the degree to which language is embedded in the environment. The more language is embedded in the environment, the larger the chance that mental simulation will be a route to comprehension. This notion is relevant for L2 learners learning the L2 in a foreign language context. In such a classroom learning context, the learner may rely primarily on books as an L2 learning method and lack in worldly, perceptual experiences needed to understand the world described by the target language (Hymes, 1972; Collentine & Freed, 2004). An interesting question is whether these learners rely on mental simulation as a route to comprehension or not.

The notion of mental simulation is central in Paper 3 in this thesis. In the next paragraph we review empirical evidence for and against mental simulation of object properties and motion in L1 and L2 comprehenders.

2.2.1 Empirical Evidence for and against Mental Simulation

In 2001 Stanfield and Zwaan designed an experimental task to examine Barsalou´s (1999) ideas about mental simulation. They investigated whether individuals mentally simulate the orientation of objects. They presented participants with sentences where object orientation was implied by the context, such as: *Harry puts the book on the shelf.* They reasoned that, typically, one would infer that the book would be in a vertical position. Following such sentences participants were presented with a picture (e.g. of a book), with the picture presented either horizontally or vertically. Participants responded (by pressing buttons) as to whether the shown object was mentioned in the preceding sentence. Stanfield and Zwaan (2001) found that participants responded faster if the orientation of the picture matched the implied orientation in the sentence as compared with mismatching sentence-picture pairs. They argued this supported the notion that participants had mentally simulated object orientation.

The orientation finding has been replicated by Engelen et al. (2011) and Zwaan and Pecher (2012). In addition, similar "match" effects have been documented for object shape (Zwaan, Stanfield & Yaxley, 2002; Engelen et al., 2011; Zwaan & Pecher, 2012; Sato et al., 2013) and color (Zwaan & Pecher, 2012). Not only lexical items, but also the sentential context seems to play a role in determining what content is mentally simulated. Bergen and Wheeler (2010) found that grammatical aspect affected mental simulation. However, it is important to note that these findings have been challenged by reported instances of null-effects (for orientation, Rommers et al., 2013; but see Zwaan, 2014) and *mismatch* advantages in the literature (e.g. Kaschak et al., 2005; Connell, 2005, 2007). For example, Connell (2007) found that comprehenders reacted faster to pictures of objects whose color mismatched with the color described in a preceding sentence in comparison with matching instances.

There are at least four accounts for conflicting support for mental simulation. First of all, it may be that (certain) simulation effects are not robust. It is important to point out that simulation effects have been derived from minimal differences in RTs (as can be seen in Zwaan & Pecher, 2012). Second, Connell (2005, 2007) argues that primary object properties (such as shape and size) cause other simulation effects than secondary properties (such as color and orientation). Thus, it may be that some (object) properties yield more stable simulations than others. Third, differences in the timing and response nature within different tasks may cause different results across experiments. Borregine and Kaschak (2006) investigated the Action-Sentence Compability Effect (ACE) in four experiments with different timings. They found that the ACE only arose when timing provided participants with the opportunity to plan their motor response while they were processing a given sentence.

Fourth, involvement of the motor system in both mental simulation and actions as button pressing may lead to null or mismatch effects. The Theory of Event Coding (Hommel et al., 2001; see Kaschak et al., 2005 for discussion) predicts that when one motor response is completed (such as running a mental simulation), features will be bound to that outcome, making it temporarily unavailable (or less available) to other responses (such as pressing buttons). Pressing buttons on a keyboard involves up- and downwards motion of fingers to a key (on a vertical axis) and coordination of left- and right finger pressing a left or right key (on a horizontal axis). Thus, if a participant needs to make mental simulations of verticality and horizontality, button presses (as in Stanfield & Zwaan´ sentence-picture verification task, 2001) to matching instances may be delayed. This would lead to faster RTs to mismatching instances (a mismatch effect) or equal RTs to both match and mismatching instances (a null effect).

Next to the mis/match debate, we may discuss which linguistic forms affect mental simulation to what degree. We know of one study that studied the role of verbs in mental simulation. Sato et al. (2013) investigated in two experiments on simulation of object shape, how big the role of Japanese verbs was in comparison with semantic context. In contrast with head-initial languages like English, in head final languages like Japanese, verbs typically occur in sentence-final position. Thus Japanese constitutes a

test case where contextual information about objects can radically change with a verb. An example where the context implies object shape is “Mother dried the kimono outdoors neatly” (the context implies a complete kimono). An example where the verb implies object shape is “I tore the kimono apart” (the verb implies a broken kimono). Sato and colleagues varied sentential context and the nature and sentence-position of verbs and measured mental simulations at different points during sentence processing. They found that comprehenders initiate mental representations early in a sentence through semantic context. However, they may rapidly modify representations if they are followed by a verb that implies a change in object shape.

We know of two studies that have investigated mental simulation in bilinguals. Vukovic and Williams (2014) found within a simulation paradigm that advanced, Dutch-English bilinguals activated task-irrelevant meanings of interlingual homophones. For example, after hearing the English sentence “On the plate in front of you/at the far end of the table, you can see a bone,” participants would see a picture showing a bean — the word for which in Dutch is “boon” /bo:n/. The subsequent picture depicted a bean that varied in size (large/small), such that it mis/matched the distance implied by the different sentence introductions. Participants were slower to reject critical items where the perceptual features matched the implied distance relationship. This suggests that bilinguals activated task-irrelevant meanings of interlingual homophones and simulated their meaning in a detailed perceptual fashion. Thus, this study indicates that during L2 processing, mental simulation in the L1 may take place.

Tomzcak and Ewert (2015) studied whether L1 speakers of Polish and English and advanced Polish learners of L2 English simulated real versus fictive motion. An example of real motion is “John runs through the forest”, an example of fictive motion is “The road runs through the forest” (Talmy, 2000). They predicted that processing of fictive motion involves simulation of physical motion and would therefore take longer than processing of static pictures and real motion for both L1 and L2 speakers (Talmy, 2013). They presented participants with prime words (e.g. a verb indicating horizontal or vertical motion of the Figure) that matched with a following Polish or English sentence respectively. They asked participants to make meaning judgments about these sentences in Polish or English and registered Yes and No answers and RTs. They found that both L1 and L2 speakers had longer RTs for fictive motion trials as compared with static pictures and real motion. They interpreted this result in favor of mental simulation of motion in both L1 and L2 speakers.

2.2.2 Theoretical Challenges

As noted above, some empirical work has provided support for mental simulation yet other studies´ results question whether mental simulation routinely occurs. Thus, important questions remain regarding if and under what circumstances individuals make mental simulations. If individuals simulate, we may ask which linguistic instances cause them to make simulations. Moreover, we may ask what happens when L2 learners comprehend L2 sentences. To what degree do they rely on abstract

symbols versus grounded simulations during comprehension?

While simulation research has focused on object properties implied by a sentence context, it is important to recognize that object properties are sometimes encoded in language itself. Different world languages offer varying ways of marking object properties linguistically. In German, for example, one does not use a general verb like ´put´ to describe placing a book on a shelf; one typically employs two verbs that mark horizontal *legen* [lay] or vertical *stellen* [stand] object orientation (Fagan, 1991; Lemmens, 2006) (e.g. *Harry stellt das Buch ins Regal* [Harry stands the book on the shelf]). Arguably, such explicit information would lead to fast, unequivocal mental simulations of object orientation. Another question that we discuss in Paper 3 is whether Spanish speakers make simulations of object size. To our knowledge, no studies have investigated yet whether comprehenders make simulations of object size. In particular, we investigate size simulation through Spanish augmentative suffixes. In Spanish large object size can be indicated by adding an augmentative suffix like *–ón*, *-azo* or *-ote* to masculine nouns, or *-ona-*, *-aza* or *–ota* to feminine nouns (Gooch, 1967). For example, *una campana* [a bell], would become *una campanota*, in case it is a large bell. Do Spanish speakers simulate large object size when they read such nouns with augmentative suffixes?

Second, we may ask to what degree the bilingual comprehender relies on abstract symbols versus grounded representations. Tomczak and Ewert´s (2015) study suggest that bilingual speakers indeed make simulations while comprehending their L2. So let us take another example. An adult learner of L2 German learns about the meaning of the German verbs, *legen* [lay] and *stellen* [stand], described above. And an adult learner of L2 Spanish learns about the meaning of Spanish augmentative suffixes described above. Will L2 German learners simulate object orientation through the novel learnt verbs? And will L2 Spanish learners simulate object size through the novel learnt suffixes? If so, does simulation happen as fast as in L1 speakers of German and Spanish? or is the process slower?

2.2.3 Research Questions (Paper 3)

In sum, an additional main research question we identify is:

Paper 3:
Does language-specific information in verbs and suffixes lead to mental simulation in L1 and L2 speakers? If so, do L1 and L2 speakers simulate with same or different speed?

2.3 L2 Learner Differences in Cognition

As mentioned in the Introduction, the specific goal of this thesis is to investigate language and perception in L2 learners in a foreign language context (e.g. Spanish learners of L2 German in Spain). These L2 learners are typically classroom learners, who learn the L2 via the means of the first

(Weinreich, 1953). However, as L2 acquisition is a highly individual process, it is important to recognize that L2 learners may differ in a number of critical aspects (Dörnyei & Skehan, 2003). An important question is whether such differences affect how L2 learners process and understand their L2. Another important question is whether differences in L2 processing and understanding affect other cognitive processes, such as learners´ memory. The chief factor commonly identified to affect cognition in L2 learners is L2 proficiency (Jarvis & Pavlenko, 2007; VanPatten, Williams & Rott, 2004; Kroll & De Groot, 2005). Two important other factors that are related to L2 proficiency are L2 usage or length, frequency and intensity of language exposure (Jarvis & Pavlenko, 2007); and motivation (N. Ellis, 2004). We discuss each of these factors in turn.

Research has shown that L2 proficiency affects various aspects of cognition, yet these effects are not clear-cut (Jarvis & Pavlenko, 2007). An important reason for ambiguity across different studies is that proficiency is measured differently. For example, some studies use years of instruction, others use length of residence, and others use various types of proficiency tests. Also, different studies look at different ranges of proficiency levels. Another important reason for ambiguity is that effects of proficiency can work differently in different areas of language acquisition and use. For example, Jarvis and Pavlenko (2007) point out that the effects of proficiency on lexical and morphological (cross linguistic) transfer often seems to be curvilinear, whereas they seem to follow more of a steady trend in areas such as word order and pronunciation (e.g. R. Ellis, 1994; Jarvis, 1998; Odlin, 1989).

Length, frequency and/or intensity of language instruction or other types of L2 exposure are often considered in studies dealing with L2 users in a foreign language context. These L2 use factors are taken to give an indication of learners´ level of language knowledge (Jarvis & Pavlenko, 2007). Length of language exposure is usually measured in the relation to the number of years of instruction a person has received in the L2 (e.g. Jarvis, 2000; Sjoholm, 1995). Frequency and intensity of language exposure are usually measured in terms of hours per day or per week of L2 instructions (e.g. Kecskes & Papp, 2000), or in terms of cumulative hours of contact they have had with the L2 (e.g. Cenoz, 2001). Research has documented effects of L2 use factors on cross linguistic transfer in word choice (Cenoz, 2001); and verb choice (Sjoholm, 1995). Kecskes and Papp (2000) found that if intensity of L2 instruction increased the better L2 learners´ writings became.

Researchers and L2 teachers will agree that motivation to learn the L2 is a determinant factor in successful L2 acquisition (Gardner, 1982; Gardner & Lambert, 1972; Dörnyei & Skehan, 2003). Motivation is defined as the learner´s orientation with regard to the goal of learning an L2 (Crookes & Schmidt, 1991). In particular, motivation concerns (i) the choice of a particular action, (ii) the persistence with it and (iii) the effort expended on it (Dornyei & Skehan, 2003). Research shows that motivation influences how often students use L2 learning strategies, how much students interact with native speakers, how much L2 input they receive, how well they do on curriculum-related achievement

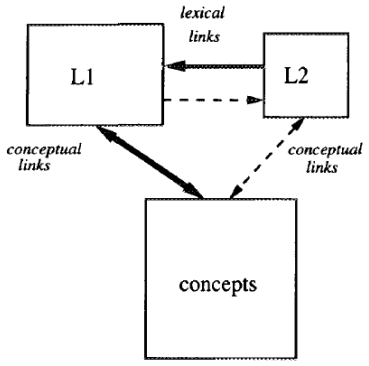
tests, how high their general proficiency level becomes, and how long they persevere and maintain L2 skills after language study is over (Gardner, 1992; Scarcella & Oxford, 1992).

Now that we have discussed the three individual difference factors of our interest in a general manner, we focus on how these factors may affect cognitive process of our interest. First, we discuss how differences may affect the processing and comprehension of (spoken or written) language. Second, we discuss that if such differences exist, whether and how differences may affect learners´ (verbal-visual) memory and mental simulation. We review empirical evidence for the notion that L2 learner differences may affect L2 learners´ comprehension and memory.

2.3.1 Empirical Evidence For L2 Learner Differences in Cognition

Several empirical studies suggest that L2 processing changes during acquisition in late or adult L2 learners. Considering words, it has been found that in early stages of L2 learning, indicating lower L2 proficiency, lexical items are processed through association with their translation equivalents in the L1. In later stages of L2 learning, indicating higher L2 proficiency, processing of L2 words is more directly conceptually mediated (Dufour & Kroll, 1995; Kroll & Stewart, 1994). This idea has been expressed in Kroll and Stewart´s (1994) Revised Hierarchical Model (RHM) on lexical and conceptual representation in bilinguals (also see Brysbaert & Duyck, 2010; Kroll et al., 2010). The model is depicted in Figure 2. In Figure 2, the L1 is represented as larger than L2 because even for fluent bilinguals, more words are known in the L1 than in the L2. Lexical associations from L2 and L1 are assumed to be stronger than those from L1 to L2, because L2 to L1 is the direction in which L2 learners first acquire the translations of new L2 words. The links between words and concepts are assumed to be stronger for L1 than for L2. There is neurological evidence for the L2 proficiency effects described by the RHM. Abutalebi, Cappa and Perani (2005) summarize neurological studies into L2 processing and conclude that increased L2 proficiency seems to be associated at the neural level with the engagement of the same network sub serving the L1, which is in line with the RHM.

Figure 2 (from Kroll & Stewart, 1994): the RHM of lexical and conceptual representation in bilingual memory.



Moving from the word to the sentence level, research has shown that L2 learners with higher levels of L2 proficiency process sentences almost as automatic as L1 speakers whereas speakers at lower levels of L2 proficiency do not (Rossi, Gugler, Friederici & Hahne, 2006; Van Hell & Tokowicz, 2010; Foucart & Frenck-Mestre, 2013). For example, Rossi et al. (2006) measured L1 Italian and L1 German speakers´ and lower and higher proficient L2 Italian and German learners´ event-related potentials (ERPs) during sentence processing. Sentences contained word category violations and morphosyntactic agreement violations. High-proficiency learners in both languages showed the same ERP components as L1 speakers for syntactic violations. In addition, the timing of processing steps was equivalent to that of L1 speakers, though some amplitude differences were present. Low-proficiency L2 learners, however, showed qualitative differences in the agreement violation characterized by an anterior negativity (LAN). A LAN is characterized by a negative-going wave that peaks around 200 milliseconds or less after the onset of a stimulus. They also showed quantitative differences reflected in a delayed P600 in every violation condition, indicating more uncertainty and problems during syntactic reanalysis. The P600 is a peak in electrical brain activity that occurs when processing grammatical errors. All in all, these results suggest that L2 processing may differ qualitatively and quantitatively at lower and higher levels of L2 proficiency.

If L2 learner differences affect L2 processing, in turn, do they affect other cognitive processes, such as memorizing? L2 processing is said to make greater demands on cognitive resources than L1 processing (Abu-Rabia, 2003; Ransdell, Arecco & Levy, 2001). At lower levels of L2 proficiency this cognitive load may be higher as compared with higher levels of L2 proficiency. It has been shown that a higher cognitive load impairs both recall and recognition (e.g. Hicks & Marsh, 2000; Naveh-Benjamin, Craik, Guez & Dori, 1998). For example, Hicks and Marsh (2000) made participants perform tasks as randomly generating numbers (1-10) or letters (A, B, C, D) or adding up sequences of numbers (1-9) during a recognition memory task. They found that memory performance, both in terms of accuracy and reaction time, was impaired in all these conditions as compared with a control group that performed no additional tasks. Considering findings on L2 processing, cognitive load and memory impairment, L2 learners with lower L2 proficiency may have worse recognition memory than L2 learners with higher L2 proficiency.

Another question one may ask is if differences in the representation and processing of one´s L2 leads to differences in L2 comprehension. In particular, in case L2 comprehension involves "mental simulation" (Barsalou, 1999; Stanfield & Zwaan, 2001) do individual differences lead to differences in mental simulation? Vukovic and Williams (2015) have shown that individual differences in spatial cognition affect comprehension in L1 speakers. They identified preferential usage of egocentric and allocentric reference frames in individuals and found in one of their experiments that only the egocentric group showed a match effect on the standard sentence-picture verification task. The allocentric participants did not produce evidence of the same effect. This indicates that individual differences may indeed affect whether one makes mental simulations. To our knowledge, no other studies have investigated

mental simulation and individual differences in L1 speakers, let alone in L2 speakers. However, it is plausible that L2 learners in a foreign language context who have greater exposure to L2 input outside the classroom, a higher degree of L2 proficiency and motivation to engage with the L2 may rely more on grounded simulations than lower proficient and less motivated learners, who are exposed to a lesser amount of L2 input.

2.3.2 Theoretical Challenges

Empirical work indicates that L2 learner differences may affect cognition. In particular, L2 proficiency plays a role, though related factors as language exposure and motivation may reveal L2 learner difference effects in cognition as well. Important questions remain in relation to differences in language comprehension and effects of language on memory. In Paper 2 and 3 of this thesis, we thus expand on previous theoretical work and address such questions.

In Paper 2 we investigate whether general L2 proficiency plays a role in visual recognition memory after verbal-visual encoding. Considering findings on L2 processing, cognitive load and memory impairment, we predict that L2 learners with lower L2 proficiency will have worse recognition memory than L2 learners with higher L2 proficiency.

Another question that we address in Paper 3 is whether individual differences in L2 learners affect if and how they make mental simulations. In particular, we investigate whether factors like L2 proficiency, language use and motivation affect mental simulation. The prediction is that the higher the (general and specific) L2 proficiency, the language use and motivation are, the faster L2 learners will process meaning and the bigger the chance that they will make mental simulations as shown by match effects.

2.3.3 Research Questions

In sum, two further main research questions we identify are:

Paper 2:
Does L2 proficiency affect verbal-visual memory?

Paper 3:
Do L2 proficiency and factors related to L2 proficiency affect meaning processing and language comprehension?

PART II:

As explained in Chapter 1: Introduction and discussed in Chapter 2: Part I, cross linguistic comparison is a critical tool for investigating language-perception interactions. The goal of this section is to introduce the linguistic (and conceptual) domains on which the empirical research in this dissertation is performed.

2.4 Cross linguistic Differences in the Description of Placement Events

A domain of investigation that has received much attention in cross linguistic approaches to language is that of motion (in space). Leonard Talmy´s typology of motion events (1975, 1985, 2000) has greatly contributed to this attention. He showed that different aspects of motion are expressed in all languages, yet different languages may show preferences to describe motion in a certain way, which are "colloquial, frequent and pervasive" (Talmy, 2000:166). A particular type of caused motion event is a "placement event". A placement event is an event where someone moves an object to another location (e.g. *Harry puts the cup on the table*). Research suggests that the basic components in placement events are: Figure (what is moved), Agent (the causer of the movement), Ground (the location where an object is placed), Causation (what triggers the placement), Motion (the act of moving), and Path (the trajectory followed by the Figure) (Talmy, 1985; Jackendoff, 1990).

Considering placement events, we can identify cross linguistic differences in the description of Motion/Path but also in descriptions of Agents and Figures. First of all, linguistic research has shown that speakers of different world languages employ different verbs to describe the Motion and Path of "putting" and "taking" actions (Kopecka & Narasimhan, 2012). Second, we find that languages vary in whether they encode gender of Agents through personal pronouns (Hellinger & Bußmann, 2001, 2002, 2003). Third, we find cross linguistic differences in descriptions of the size of Figure objects (Savickiene & Dressler, 2007). These differences make placement events an interesting domain to investigate language-perception interaction from multiple perspectives. In particular, a comparison of the German and Spanish language provides an interesting study case as these languages show differences in their colloquial descriptions of Motion/Path, Agents and Figure.

In the following we discuss cross linguistic differences in German and Spanish in the description of Motion/Path, Agents and Figure. We focus on linguistic instances of Standard German and Peninsular Spanish. First, we describe how German verbs mark the spatial orientation of objects whereas Spanish placement verbs do not. Second, we discuss a case where Spanish personal pronouns indicate the gender of agents whereas German pronouns do not. Third, we outline how augmentative suffixes can provide information about the size of objects in Spanish whereas such morphological marking is not possible in German. If applicable we provide information on alternative meanings of the linguistic

instances. If available we provide information on how and how frequently speakers use them in daily life.

2.4.1 Space marked by Placement Verbs

The action of putting (Motion/Path) is central in any given placement event. Interestingly in German and Spanish (and in other Germanic languages as Dutch and Romance languages as French) this action is described by means of different verbs. These verbs differ as to whether they provide spatial information about the orientation of the placed object – the Figure – with respect to the Ground.

In German, one employs a set of semi-obligatory placement verbs that mark the position of objects or persons in relation to the Ground (Fagan, 1991; Lemmens, 2006). More specifically, the positional verbs *stellen* [stand] and *legen* [lay]) indicate whether the final position of the moved entity is either vertical (*stellen*) or horizontal (*legen*) in relation to the Ground. In contrast, Spanish typically employs a single placement verb *poner* [put] or *dejar* [leave in place]) that does not indicate a certain position of an object or person in relation to the Ground (Ibarretxe-Antuñano, 2012; Cadierno et al., 2016).

It has been posed that verbs of "putting" and "taking" are amongst the most frequent, basic verbs in a language and that they are amongst the earliest verbs learned by children (Levinson, 2012). We know of no corpus studies into how frequent the particular German and Spanish instances described above appear in spoken and written language. However, Lemmens (2006) has conducted a corpus study on the Dutch placement verbs *zetten* [stand] and *leggen* [lay] and *steken/stoppen* [stick (into)]. He found 13814 extractions of these placement verbs within a (mainly written) corpus of 24.9 million words.

2.4.2 Gender marked by Personal Pronouns

Another relevant element in placement events is the Agent performing the action. There may even be multiple agents performing the action as in: *They put their glasses on the table*. A human agent naturally can be recognized by aspects such as body shape, height and weight. Moreover, an agent may display gender characteristics or aspects that are considered either as feminine or masculine. Gender is expressed differently in nouns and personal pronouns in different world languages (Hellinger & Bußmann, 2001, 2002, 2003). For German and Spanish, we find a cross linguistic difference in the linguistic expression of gender. In particular, concerning the personal pronoun "They", Spanish marks a masculine and a feminine "They" (i.e., *ellos* [they-masculine] and *ellas* [they-feminine]) whereas German does not make such a distinction (though German does mark gender for the third person singular).

In Spanish, a group with women is referred to as *ellas* [they-feminine]; a group with men only, or both men and women, is referred to as *ellos* [they-masculine] (Alarcos Llorach, 1994; Bosque & Demonte, 1999; Bosque, 2010; Álvarez Martínez, 1989). We note that Spanish is a "pronoun-dropping"

language. This means that pronouns may be omitted from speech or writing unless they are needed for purposes of emphasis or contrast (De Bruyne, 1993; Alarcos Llorach, 1994; Butt & Benjamin, 1998).

In German, a group of two or more anonymous persons can be referred to by the personal pronoun *sie* [they]. We note that the form *sie* is also used to describe a single female individual "she" or, when written with a capital S, "Sie" may mean "you" expressed in a polite manner (Drosdowski, 1995; Cartagena & Gauger, 1989).

We know of no corpus studies into how frequent the particular German and Spanish instances described above appear in spoken and written language. However, it has been posed that in any language personal pronouns constitute a basic and culturally significant lexical field. They are needed to communicate about the self and others and they are used to identify people as members of various groups (Hellinger & Bußmann, 2001, 2002, 2003).

2.4.3 Size marked by Augmentative Suffixes

Another element of importance in placement events is the Figure-object that is being placed. Similar to Agents, objects have different properties, such as a typical orientation, shape, color, texture and size. We find that German and Spanish provide different means to describe the size of an object. In particular, in Spanish we can provide information about large object size by adding an augmentative suffix to nouns whereas German does not offer this option.

In Spanish, large object size can be indicated by adding an augmentative suffix like *–ón, -azo* or *-ote* to masculine nouns or *-ona-, -aza* or *–ota* to feminine nouns (Gooch, 1967). For example, *un libro* [a book], would become *un librote,* in case it is a large, heavy book. We note that Spanish augmentative suffixes are primarily used to denote large size, but they have a secondary function of adding an emotional tone to a given word (Butt & Benjamin, 2005; Hualde et al. 2010; RAE, 2009). This emotional tone is mostly associated with either admiration or a pejorative idea of clumsiness, unpleasantness, awkwardness or excess, as in *se me ha pegado un catarrazo* (I´ve caught one heck of a cold; *un catarro*= a cold).

In German, one cannot express large object size by adding a suffix to a noun (Lohde, 2006; Korecky-Kröll & Dressler, 2007). If one would want to describe something that is large in size or describe reinforcement/intensification, one would need to add nouns such as *Riese(n)-*, *Bombe(n)-*, *Spitze(n),* with positive connotations; or *Hölle(n)-*, *Heide(n)-*, *Pfund(s)-*, that have negative connotations, to the base noun one wishes to alter. For example, a big success (*Erfolg*), would become a *Riesenerfolg* (Lohde, 2006). We note that a positive or negative connotation of augmentative constructions is of course also dependent on the textual context.

We know of no corpus studies into how frequent Spanish augmentatives appear in spoken and written language (but see Carranza & Seguí, 2005 for a frequency indication for Argentinean Spanish). However, we know that augmentatives are related to pejorative and diminutive morphology. Diminutives emerge as the earliest category of derivational morphology in nearly all languages and of morphology as a whole in many languages (Savickiene & Dressler, 2007). It has been reported in Spanish dictionaries and manuals that the forms and frequencies of Spanish augmentative suffixes (as well as diminutives) differ from place to place in the Spanish-speaking world. They seem to appear rather in spoken than in formal written Spanish and they are said to be more common in women´s speech than men´s in some regions (Butt & Benjamin, 2005; RAE, 2009).

2.4.4 Linguistic Saliency

In sum, we have introduced the linguistic and conceptual domains on which the empirical research in this dissertation is performed. In particular, we have discussed linguistic differences in the expression of Motion/Path, Agents and Figure in placement events between Standard German and Peninsular Spanish. We have described how German verbs mark the spatial orientation of objects, whereas Spanish placement verbs do not; how Spanish personal pronouns indicate the gender of multiple agents whereas German pronouns do not; and how Spanish augmentative suffixes can provide information about the size of objects, whereas such morphological marking is not possible in German. We found no corpus studies that reveal how frequent our forms of interest appear in spoken and written German and Spanish. However, we found claims by several authors that all linguistic forms of interest constitute basic and significant lexical fields.

The general perceived strength of linguistic forms and thereby the concepts they describe is commonly referred to as their salience (Ellis & Cadierno, 2009; Ellis, 2008; Ellis, 2006; Goldschneider & DeKeyser, 2001). Linguistic forms may be more or less salient and attract more or less attention (Talmy, 2000; Ellis, 2008). For example, in the sentence ‘She starts school today´ the word ´today´ is a stronger psychophysical form in the input than is the morpheme ´-s´ marking third person singular present tense. While both provide cues to present time, ´today´ is much more likely to be perceived, and ´-s´ can thus become overshadowed and blocked. Several factors determine the saliency of a linguistic form (Talmy, 2008; Ellis, 2008). For example, frequency of occurrence of a linguistic form in language plays an important role (Berman & Slobin, 1994; Larsen-Freeman, 1976; Talmy, 2000; Ellis, 2008). Berman and Slobin (1994: 640) have posed that “(…) frequent use of forms directs attention to their functions, perhaps even making those functions (semantic and discursive) especially salient on the conceptual level. That is, by accessing a form frequently, one is also directed to the conceptual content expressed by that form.”

Another important factor that determines the saliency of a particular linguistic form is word class. For example, relevant to the placement verbs discussed in § 2.4.1, Leonard Talmy (2000; 2008) argued that, other things being equal (such as a form´s degree of stress or position in the sentence) semantic

components expressed by the main verb root are backgrounded and thus less salient. If expressed elsewhere in the sentence they would be foregrounded. For example, object orientation is foregrounded in sentence (1a) due to its description in an adverb phrase. In sentence (1b) however it is a more incidental piece of background information conflated within the main verb.

(1a) I put the binoculars on the shelf in a vertical manner.
(1b) I stand the binoculars on the shelf.
(1c) I put the binoculars on the shelf.

Talmy (2000) poses that speakers tend to opt for the expression of a concept over its omission more often when it can be referred to in a backgrounded way. Also, it tends to be stylistically more colloquial, or less awkward, where it can be backgrounded than when it must be foregrounded. He further argues that where a concept can be backgrounded, its informational content can be included in a sentence with apparently low cognitive cost – in particular, without much additional speaker effort or hearer attention. For example, the informational content of sentence (1b) is similar to (1c), yet sentence (1b) adds that the binoculars are in a vertical position while it can be expressed with as little speaker effort or hearer attention as sentence (1c). Following these notions we may predict that the placement verbs discussed in § 2.4.1 may have limited salience for language users when presented in a sentence such as (1b).

In an article that discusses aspects of attention in language, Talmy (2008) describes how a multitude of factors with differing degrees of salience combine and interact in establishing attentional effects. Some of these factors involve formal aspects of language (properties of the morphemes, syntax, phonology). He discusses attentional effects resulting from combining factors for "agency", which is relevant to the personal pronouns discussed in § 2.4.2. He argues that attention on agency incrementally increases by the successive addition of factors in the following series of otherwise comparable sentences. These sentences are all taken to refer to the same scene in which a group of friends –the Agents- hand a cup of wine to another as they sit around a table.

(2a) The cup went around the table.
(2b) The cup was passed around the table by them.
(2c) They passed the cup around the table.
(2d) The friends passed the cup around the table.

In (2a) a minimal backgrounded sense of agency is pragmatically inferable from the context, but it is not specifically represented by the linguistic forms themselves. In (2b) the transitive verb *pass* makes agency slightly more salient because it includes indirect reference to an agent. However, in (2b) "a sharp rise in attention on the agent occurs when it is explicitly referred to by an overt pronoun" (Talmy, 2008:35). In (2c) the agency is further foregrounded by the occurrence of a pronoun as subject in initial position. Finally, replacement of the pronoun by a full lexical noun as in (2d) foregrounds the agent to

the greatest degree. Following these notions we may predict that the personal pronouns discussed in § 2.4.2 will at least attract some of language users´ attention to Agents when presented in a sentence like (2c).

Talmy (2008) also points out that some factors may raise or lower attention to a referent because of its properties. One such factor deals with referential divergence from norms and is relevant to the Spanish augmentative suffixes discussed in § 2.4.3. Talmy (2008) poses that a referent´s divergence from certain norms tends to foreground it. Such norms and deviations from them include ordinariness versus unusualness. For example, a more unusual referent, such as a huge, heavy object (e.g. a huge, heavy book, flashlight or lipstick) tends to attract greater attention than a more ordinary referent, such as an object in prototypical size (e.g. a prototypically sized book, flashlight or lipstick). Following this notion we may expect that Spanish augmentative suffixes and the large object size they indicate will attract language users´ attention. However, considering more formal aspects of language, augmentative suffixes are bound morphemes that are bound to the (object) nouns they alter. Ellis (2008: 380) has pointed out that " (…) grammatical function words and bound inflections tend to be short and low in stress (…) with the result that these cues are difficult to perceive". Thus, a combination of referent properties and more formal aspects of language may lead augmentative suffixes to attract language users´ attention in lesser a degree as compared with considering referent properties alone.

All in all, we have discussed that different factors may contribute to situations where placement verbs have only limited salience for language users; personal pronouns have at least some degree of salience for language users and we have found conflicting notions that indicate that Spanish augmentative suffixes may either be quite salient or less salient to language users.

2.4.5 Teaching Cross Linguistic Differences

It is known that linguistic forms with low salience tend to be less readily learned by L2 learners (Ellis, 2004; 2008). In addition, it has been argued that forms that are similar to L2 learners´ L1 will be simple to learn whereas forms that are different will be difficult to acquire (e.g. Lado, 1957). These notions constitute a prime motivation for explicit L2 instruction on cross linguistic differences (with limited salience). As the focus of this thesis is the development of psycholinguistic theory on language and perception and not the development of L2 teaching materials, we do not provide an exhaustive overview on possible L2 teaching methods and materials and their merits and disadvantages respectively. Norris and Ortega (2001) have provided a meta-analysis of a large number of studies on the effectiveness of different L2 teaching methods. Based on their analysis they conclude that more important than the type of instruction or material used is the mere fact that L2 instruction takes place. They note that instructional techniques that emphasize meaning are more commensurate with what psychologists know about how the brain internalizes new knowledge (Norris & Ortega, 2001).

For purposes of conducting the studies presented within this thesis we developed class-fronted L2 instructions on the meaning of the German verbs *legen* [lay] and *stellen* [stand] (Appendix 4) and Spanish augmentative suffixes (Appendix 5) to facilitate their learning. We found that these forms are not typically taught in text books used by the *Instituto Cervantes* and the *Goethe Institut*, whose L2 learners we aimed to recruit for our studies.We did not develop an L2 instruction on the meaning of the Spanish personal pronouns *ellos* [they-masculine] and *ellas* [they-feminine] because we found that these forms are typically introduced in text books used by *Instituto Cervantes* at beginners level.

The instructions we developed can be classified as "form-focused" (FF). Rod Ellis (2001:2) defines FF instruction as "any planned or incidental instructional activity that is intended to induce language learners to pay attention to linguistic form". The term is used to describe both teaching approaches based on artificial syllabi as well as on more communicative approaches, where attention to form arises out of activities that are primarily meaning-focused (e.g. Long & Robinson, 1995). FF instruction is usually contrasted with focus-on-meaning (FoM) instruction. FoM instruction involves exposing learners to rich input and meaningful use of the L2 in context, which is intended to lead to incidental L2 acquisition (Norris & Ortega, 2001). This instruction type is in line with Krashen´s Natural Appraoch, content-based L2 instruction and immersion programmes (R. Ellis, 1994). FF instruction comprises two subcategories, namely focus-on-formS (FoFS) and focus-on-form (FoF) instruction. FoFS instruction is the traditional approach to grammar teaching, where language is treated as an object to be studied and learners are seen as students rather than users of the language (Ellis, 2001). FoF instruction involves strategies that draw learners´ attention to the form or properties of target structures within a meaningful context. This is done for forms that are potentially difficult and likely to be used or needed in future communication (Spada & Lightbown, 1993).

Our instruction can be further classified as FoF instruction. Following Terrell (1991) we see instruction as a way of increasing the salience of forms by first pointing them out and explaining their structure and that it is necessary to provide learners with meaningful input that contains many instances of the critical form-meaning relationship. This view is also in line with "processing instruction". This is a well-known type of L2 instruction where one provides explicit information about the target structure, explicit information about processing strategies and structured input activities (VanPatten & Cadierno, 1993a, 1993b; Cadierno, 1995).

Both our instruction on German placement verbs and Spanish augmentative suffixes went as follows. First, we presented learners with an 18-item exercise to find out what they already knew about critical forms (pretest). In the exercise, L2 German students indicated whether a given German verb (6 versions of *legen* [lay], 6 versions of *stellen* [stand] and 6 other verbs) indicated horizontality, verticality or no orientation. L2 Spanish students indicated whether a given Spanish noun (6 diminutive, 6 normal, 6 augmentative nouns) indicated small, normal or large size. Second, we presented L2 learners with the critical linguistic forms and explained their meaning and structure (see

Appendix 4 and 5). After this introduction we provided a meaningful example where after the teacher asked a comprehension question. Finally, L2 learners completed the same 18-item exercise as before the instruction with items in a different, randomized order (posttest). This was not only an exercise but also a test on whether learners had correctly understood the L2 instruction, with scores that could vary from 0 until 100% correct.

Chapter 3: Methodology

The aim of this chapter is to describe the samples and methodologies used to investigate the questions posed within each paper in this dissertation. The studies in this thesis are quantitative. Quantitative research can be defined as research in which there is quantification of data and numerical analyses. The studies in this thesis are also experimental, which means we manipulate variables to test hypotheses (Mackey & Gass, 2005: 363). First, I describe the participants that represent the populations we investigated. Second, I discuss the different experimental tasks we employed and measures we took to test our hypotheses. I also describe the setting wherein we tested our participants. Finally, I discuss how we treated the data and the statistical models we build to determine the validity of our hypotheses.

3.1 Participants

All participants were paid a nominal fee for partaking in the experiments and were informed about the general goal of the study before participation. We used a non-random sampling technique relying on available subjects (Babbie, 2015). This means not all individuals in the population had an equal chance of being selected. We mainly recruited university students between 18-35 years old in order to cover our target population of (younger) adults. The reader will discover in the individual papers that a handful of older participants participated in the experiments. In case their data did not differ significantly from their younger fellows, their data were included in analyses. We aimed to recruit as many men as women. However, in some of the papers we report unequal numbers of male/female participants.

3.1.1 L1 speakers

In all three papers we present data of L1 speakers of German and Spanish. All L1 German speakers were recruited at the University of Bremen (Germany) in June 2014, December 2014 and March 2015. All the L1 Spanish speakers were recruited at the University of Seville (Spain) in May 2014 and March 2015. Students were recruited through advertisement on the university Blackboard system, in university canteens and other public spaces and through professors and colleagues at the relevant universities. The university students had diverse study backgrounds, ranging from Psychology, Law, Business, English and Geography to Mathematics. Almost all students spoke English as an L2 and had limited knowledge of other languages. However, we ensured none of the L1 German speakers spoke Spanish and none of the L1 Spanish speakers spoke German.

3.1.2 L2 learners

In paper 2 and 3 we present data of German learners of L2 Spanish and of Spanish learners of L2 German. The German learners of L2 Spanish were recruited in Germany at the following locations. In May 2015, we recruited L2 learners in Bremen through *Instituto Cervantes*, which reached out to

other institutes in Bremen that gave L2 Spanish courses. The Cervantes Institute, a government agency, is with 54 centers in 20 different countries, the largest organization in the world responsible for promoting the study and the teaching of Spanish language and culture. In September 2015, we recruited L2 learners in Berlin, through *Instituto Cervantes*. As we aimed to find more participants than *Instituto Cervantes* could provide we recruited further participants through the language center of *Humboldt Universität zu Berlin* and *Volkshochschule Berlin Mitte*. The last number of L2 learners was recruited through the language center and the Spanish department at *Westfälische Wilhelms-Universität Münster* in January 2016. All participating institutions placed students at different levels of proficiency following the Common European Framework of Reference (CEFR). CEFR is a framework that has been established by the Council of Europe as a way to standardize the language learning achievements of language learners across Europe (Council of Europe, 2001). The CEFR has six levels of assessment: A1 and A2 (basic user), B1 and B2 (intermediate user) and C1 and C2 (proficient user) (see Appendix 3).

The Spanish learners of L2 German were recruited in Spain at the following locations. In October 2016, we recruited L2 learners in Granada, through the *Goethe Institut*. The Goethe Institute is a non-profit German cultural association operational worldwide with 159 institutes, promoting the study of the German language abroad and encouraging international cultural exchange and relations. As we aimed to find more participants than *Goethe Institut* could provide we recruited further participants through the German department of the *Universidad de Granada*. In November 2016, we tested more L2 learners in Seville, through the German department of the *Universidad de Sevilla*, which also reached out to other institutes in and around Seville that provided L2 German courses. All participating institutions placed students at different levels of proficiency following the CEFR.

Both German and Spanish L2 students varied in terms of their language proficiency, age of acquisition, in whether they stayed in a country where the L2 is spoken and various other aspects. This was registered by their answers to an adapted version of the Language Background Questionnaire (LBQ) (Gullberg & Indefrey, 2003) that was translated into German (Appendix 1) and Spanish (Appendix 2). Through this questionnaire, it was also confirmed that students learned their L2 in adulthood (on average, with 18 years old). In addition, almost all students spoke English as an L2 and had limited knowledge of other languages.

3.1.3 Sample Size

The number of participants in a study, or the size of a sample, is of great importance in quantitative research. The main reason is that sample size affects the statistical power of a given test and the related effect size. The power of a test is the probability that a given test will find an effect assuming that one exists in the population (Field, 2013, p.69). An effect size is an objective and (usually) standardized measure of the magnitude of an observed effect (Field, 2013, p.79). An effect size measure relevant to

analyses in thesis is eta squared or partial eta squared (Levine & Hullett, 2002). In the papers in this thesis, if possible, we based our sample sizes on sample sizes used in previous research. We verified whether we had achieved adequate power post-hoc. The sample sizes for each of the three papers in this thesis are reported in Table 1 below.

3.2 Tasks and Test Setting

3.2.1 Experimental Tasks and Measures

We employed six different experimental tasks in the studies reported in the different papers. Five of those tasks were assumed to measure non-linguistic behavior, including mental imagery, categorizing and memorizing. An additional task involved asking participants to provide linguistic descriptions of picture stimuli. See Table 1 for a schematic overview of different tasks and measures used in the three papers.

In Paper 1, we employed two types of categorization tasks: categorization by forced choice and similarity judgment. In addition, we gave participants a linguistic description task. As explained in paragraph 2.2.1, categorization by forced choice involves choosing two out three pictures that are most similar. Similarity judgment involves rating how similar two pictures are on a scale. The linguistic description task involved giving written descriptions of pictures. In the forced choice task, we registered participants´ choices; in the judgment task, we measured their judgments on a scale from 1-7. We also registered the linguistic descriptions with which the participants described a set of pictures.

Table 1 : Overview of experimental tasks, measures, number of participants and statistical tests in the four papers in this thesis.

	Experimental task	**Measure(s)**	**N Participants**	**Statistical Test**
Paper 1	Categorization by Forced Choice	Categorization Choice	L1 German: 22 L1 Spanish: 23	T-Test ANOVA
	Similarity Judgment	Rating (scale 1-7)		
	Linguistic Description	Linguistic Description		
Paper 2	Match Task	Yes/No Response Reaction time	L1 German: 27 L2 Spanish: 27 L2 German: 123 L2 Spanish: 141	ANOVA
	Picture recognition	Yes/No Response Reaction Time		
Paper 3	Sentence-picture verification	Yes/No Response Reaction Time	L1 German: 30 L2 Spanish: 34 L2 German: 122 L2 Spanish: 100	Linear mixed effect modelling

In Paper 2, we gave participants a match task and a picture recognition task. The match task involved judging whether a given sentence matched with two pictures, Yes or No. As explained in paragraph 2.2.1, the picture recognition task involved judging whether a given picture was identical to the picture(s) presented before, Yes or No. For both tasks, we measured a) whether participants gave a Yes or No response to the questions "Does this sentence match with the pictures above?" and "Is this picture identical to the picture(s) you just saw?" ; and b) their reaction time in milliseconds. We only analyzed responses to the picture recognition task. Yes or No responses indicated whether participant had correctly or falsely recognized identical and non-identical pictures. Again, higher reaction times (thus slower reactions) were assumed to reflect higher cognitive load.

In Paper 3 we employed a sentence-picture verification task. As explained in paragraph 2.1.1, the aim of this task was to measure whether participants reacting faster to *matching* sentence-picture pairs as compared with *mismatch* sentence-picture pairs. We presented participants with sentences, followed by black-and-white drawings of objects. We measured a) whether participants gave a Yes or No response to the question "Was this object mentioned in the sentence before?"; and b) their reaction time in milliseconds. Yes or No responses indicated how they had verified whether a picture was mentioned in a preceding sentence. Higher reaction times (thus slower reactions) were assumed to reflect higher cognitive load.

3.2.2 Measures of L2 Differences

We took both general (LBQ, Gullberg & Indefrey, 2003) and specific measures of differences in L2 learners. These were designed to be comparable across German and Spanish. Here we discuss measures related to L2 learner differences that we considered as most relevant to investigate. These are measures related to the factors L2 proficiency, language usage and motivation (as described in 2.3).

Considering L2 proficiency, we took two measures assumed to reflect general (Paper 2 and 3) and specific (Paper 3) L2 proficiency. In the LBQ L2 learners reported their general level of proficiency in terms of the CEFR (see Thompson, 2015 on the validity of this method). We asked learners to report the level at which they were currently studying. If they did not currently study we asked them to report the last level attained. As a more specific measure of language proficiency, we were interested in how well L2 learners knew critical language-specific verbs and suffixes. Therefore, we designed two 18-item tests to test learners´ knowledge of target forms (see Appendix 4 and 5). In the German version of the test, L2 German students indicated whether a given German verb (6 versions of *legen* [lay], 6 versions of *stellen* [stand] and 6 other verbs) indicated horizontality, verticality or no orientation. In the Spanish version of the test, L2 Spanish students indicated whether a given Spanish noun (6 diminutive, 6 normal, 6 augmentative nouns) indicated small, normal or large size. Test scores could vary from 0-100% correct.

Considering language usage, L2 learners reported two relevant measures in the LBQ (used in analyses Paper 3). First, they reported the number of years of instruction they received in the L2. This measure took into account both formal instruction (in a classroom context) as well informal learning (by interacting with native speakers on a frequent basis). Frequency and intensity of language exposure was measured in terms of reading, listening, speaking and writing in the L2 in hours per day.

Motivation (used in analyses Paper 3) was reported in the LBQ by rating three statements on a scale from 1-5 (e.g. "I like speaking [L2]; "I feel secure using [L2]"; "I think it is important to know [L2]").

3.2.3 Test Setting

The experimental tasks we employed in the different studies were computer-based. Therefore, all participants were tested in a quiet computer room at the different testing locations in Germany and Spain. After completing the experimental tasks all subjects filled in the LBQ. Students were tested in groups of 5-30 people. The principle investigator, which is the author of this thesis, led experimental sessions in almost all cases. Sessions were led in one of the relevant languages, German or Spanish, which are second languages of the principle investigator. L1 speakers received instructions in their L1; L2 learners in their L2 (with possibility to ask questions in their L1, in case they did not understand L2 instructions). In addition, L2 students studied word lists (Appendix 6 and 7) with 15 critical and 15 distractor words before embarking on the computer experiments. On several occasions, the principal investigator was aided by a student assistant that was a native speaker of the relevant experiment language. Half of the L2 participants in Berlin were tested by a student assistant without the presence of the principle investigator. The latter factor was treated as a factor in (unreported) exploratory statistical analyses for Paper 3, but was not found to be significant.

We offered participants to choose from several 90 minute slots, either in the morning, afternoon or early evening for their participation. Almost all L2 participants completed two different experimental tasks. To reduce the chance of any effects between experiments, the order of administration of different experimental tasks was counterbalanced. Experimental order was treated as a factor in (unreported) exploratory statistical analyses in Paper 2 and Paper 3, but was not found to be significant.

3.3 Data Treatment and Analysis

3.3.1 Data Treatment

When modeling data statistically one needs to be aware of bias (Field, 2013: 163). When we speak of bias, we mean that something is not evaluated in an objective way: there are other aspects affecting the conclusion. An important source of bias in linear models is "outliers". An outlier is a score that is very different from the rest of the data. Such an outlier can bias a parameter such as the mean or median. It

can have an even greater influence on the error associated with that parameter. If the sum of squared errors is biased, so are the standard error and the confidence interval associated with the parameter. In addition, most test statistics are based on sum of squares so these will be biased too by outliers. Another important source of bias is "violation of assumptions". Most statistical models require that your data meets a number of assumptions, such as that your data is normally distributed. If this is not the case, a resulting statistical model may be biased.

There are essentially four ways of reducing bias. The first is to trim the data that is to delete a certain amount of extreme scores. Often, trimming follows one of two rules: (1) a percentage based rule (e.g. delete 10% of highest and lowest scores); or (2) a standard deviation based rule (e.g. delete scores further away than two standard deviations from the mean). Second, one may "winsorize" data, which means one substitutes outliers with the highest value that is not an outlier. Third, one may analyze with robust methods, which typically involves a technique known as bootstrapping. Robust tests estimate statistics that are reliable, even when assumptions of the statistic are not met. Finally, one may transform the data. This involves applying a mathematical function (e.g. log transformation) to scores to try to correct any problems with them.

In this thesis, we employ all four methods of reducing bias described above. In each paper, the method was chosen with consideration of data treatment in previous related work. This makes our work comparable to other studies within the field. Details of trimming procedures are discussed in each individual paper. For all statistical models reported in this thesis, model assumptions were met, unless otherwise reported.

3.3.2 Statistical Models

There is currently great critical interest regarding the quality of quantitative data analysis. Many researchers suggest that traditional null-hypothesis significance testing (NHST) does not suffice anymore (Cumming, 2014; Kline, 2004; Rouder et al., 2009). NHST relies on fitting a model to the data and then evaluating the probability of this model, given the assumption that no effect exists. We are aware of arguments in favor of Bayesian analyses (e.g., Kruschke, 2010). These analyses take base rates into account when determining whether any given data supports a hypothesis or not. In this thesis, results were analyzed using traditional NHST. The main reason for this was to maintain comparability to previous research. We do report effect sizes and confidence intervals to help evaluate findings without dogmatic reliance on p-values (Cumming, 2012).

The statistical test one chooses to analyze data depends on one´s research questions and experimental design. In the experiments in this thesis, there was variability in both subjects and items. Traditionally, such designs are analyzed using (mixed-model) repeated measures analysis of variance (ANOVA) (or a T-test, in case one wants to compare only two group means). ANOVA is used to analyze differences

within and between group means. It does so by comparing the ratio of systematic variance to unsystematic variance (or error) in an experimental study. The ratio of these variances is the *F*-ratio. This test statistic comes with a p-value with which one assesses whether the mean of several groups are equal or not. If the difference between the group means is large enough, then the resulting model will be a better fit of the data than the grand mean model, which poses there is no relationship between the predictor variable and the outcome. ANOVA is considered to be a "robust" test, which means it still gives accurate information when assumptions are being violated (but see Field, 2013: 444).

There is a growing trend to analyze experiments with variability in subjects and items using linear mixed-effect models (e.g. Barr, Levy, Scheepers & Tily, 2013). These models are also referred to as "hierarchical linear models" (Raudenbusch & Bryk, 2002), "random coefficients models" (Longford, 1993) or "multilevel models" (Luke, 2004). There are empirical, statistical and theoretical reasons to employ mixed-effects models instead of ANOVA (see Luke, 2004:7). The main theoretical argument in favor of a mixed-effect model is that if the phenomenon you are studying is multilevel in nature, then the analytic technique should also be multilevel. An example of a multilevel structure is students being nested in classrooms. It is assumed that students in the same classroom will be more alike than students in a different class. Another example is time being nested within a person. This applies to experimental designs where participants are being tested multiple times. Here it is assumed that time may affect participants´ behavior.

The goal of a multilevel model is to predict values of some dependent variable as a function of predictor variables at more than one level. Multilevel models are also called "mixed models", as they are always made up of both *fixed* and *random* effects. For example, imagine that a researcher wants to compare two teaching methods used in nine different schools in a city. S/he would want to generalize inferences about the teaching methods´ effects to the population of the schools in this area. Thus, when modeling the data, s/he would include teaching method as a fixed and school as a random factor. How well fixed and random variables describe the data is assessed by means of comparing models (Luke, 2004). One may compare maximum likelihood (ML) or restricted maximum likelihood (REML); Akaike Information Criterion (AIC) or Schwarz´s Bayesian Information Criterion (BIC) (Akaike, 1987; Schwarz, 1978).

In the three different papers, we employ two main types of statistical tests (see Table 1) based on the research questions asked. In Paper 1 and 2 we analyzed data with ANOVAs. In Paper 3 data was analyzed data using linear mixed-effects models.

3.3.3 Selection of L2 Learner Differences

In Paper 2 and 3 we were interested to analyze whether L2 learner differences affected L2 learner´s behavior in the tasks described in 3.2.1. These were tasks examining recognition memory (Paper 2) and meaning processing and mental simulation (Paper 3).

In Paper 2, we were interested in effects of general L2 proficiency. Therefore, we used the CEFR levels that L2 learners reported in the LBQ. We used these measures to create a group of beginning (A2 level, intermediate (A2-B1 and B1) and advanced L2 learners (B2 level or higher). These groups were used in the analyses in Paper 2.

In Paper 3, we were interested in effects of general L2 proficiency, knowledge of critical linguistic forms, language usage and motivation. Exploratory analyses revealed that the CEFR proficiency indication and self-rated proficiency rated on a scale of 1-5 did not reliably affect RTs. Instead of these measures, we considered number of years of L2 instruction as indicator of L2 proficiency. Knowledge of critical linguistic forms was measured by a pre- and posttest taken before and after L2 instruction on the critical forms. We also considered language usage as indicated by daily L2 use in hours and we considered motivation, which was measured by computing an average based on the ranking on the three statements described in 3.2.2.

4. Summary of Findings

The aim of this chapter is to give an overview of the findings within the different papers. I summarize the findings per paper and provide and describe a table where I present the results for each German-Spanish cross linguistic difference (in the expression of space, gender, size) that we investigated. This overview will be helpful in the discussion of the results in Chapter 5.

4.1 Summary of Findings per Paper

In Paper 1 we investigated L1 speakers of German and Spanish. We investigated whether language categories that differ across German and Spanish affected speakers´ categorization of placement scenes in two categorization tasks where participants are not presented with or encouraged to produce language. The relevant language categories concerned German-Spanish cross linguistic differences in the expression of space and gender in placement scenes. The main hypotheses were:

1A. German speakers will give placement scenes with same object position higher similarity ratings than Spanish speakers. Spanish speakers will give scenes where the gender of agents is the same higher similarity ratings than German speakers.

2A. German speakers will categorize placement scenes on the basis of similarity in object position whereas Spanish speakers will categorize on the basis of similarity of the gender of agents.

After participants completed the categorization tasks, we examined whether they described placement scenes following our linguistic predictions. In particular, we examined whether German speakers described placement scenes by means of the pronoun *sie* [they] and the verbs *legen* [lay] and *stellen* [stand] and whether Spanish speakers described placement scenes by means of the pronouns *ellos* [they-masculine] and *ellas* [they-feminine] and the verbs *poner* [put] or *dejar* [leave in a place].

The main results were as follows. First, we examined whether German and Spanish speakers´ similarity judgments and categorization choices differed significantly. In particular, we examined how high they rated the similarity of placement scenes where object orientation was similar and scenes where the gender of Agents was the same on a scale from 1-7. Also, we counted how often German and Spanish speakers´ forced categorization choices were based on similarity in object position or on similarity in gender of Agents. We found that speakers´ similarity ratings for space and gender aspects in placement scenes did not follow distinctions that are made in their language. In other words, German and Spanish speakers´ similarity judgments and forced categorization choices did not differ significantly and Hypothesis 1A and 2A were not supported. In the next chapter we elaborate on possible explanations for why we did not find an effect. On the one hand, speakers may have not used language while completing the tasks (though language use was not prohibited by performing a task such as counting

backwards during categorization). Therefore, cognitive differences guided by cross linguistic differences did not occur. On the other hand, speakers may have used other language than the language predicted during the tasks. Such alternative descriptions may have resulted in similar indication of object position and gender of agents across languages.

We also examined how the same German and Spanish speakers described placement scenes in a description task given after the categorization tasks. We found that German speakers used the predicted pronoun in 19.3 % of the cases and the predicted verbs in 39.8 % of the cases. Spanish speakers used the predicted pronouns in 63.0% of the cases and the predicted verbs in 17.9 % of the cases. Through qualitative and quantitative analysis of their alternative descriptions we found that German and Spanish speakers indicated object position and gender of agents equally often in their descriptions of placement scenes. German speakers indicated object position 48.9 % of the cases and Spanish speakers did this in 54.3 % of the cases. German speakers indicated the gender of agents in 86.9% of the cases and Spanish speakers did this in 86.4% of the cases. Thus, in case these descriptions were employed as a mental strategy while completing the categorization tasks it is not surprising we did not find differences in categorization preferences within this study.

In Paper 2 we investigated both L1 and L2 speakers of German and Spanish. We investigated whether language affected their recognition memory in a computer task. The relevant language concerned German-Spanish cross linguistic differences related to the expression of space and gender in placement scenes. These were the same differences as in Paper 1. The following hypotheses stood central:

1B. L1 German speakers will have a better recognition memory for object orientation than L1 Spanish speakers and conversely, L1 Spanish speakers will have better recognition memory for facial gender of agents.

2B. When Spanish learners of L2 German and German learners of L2 Spanish perform the task in their L2, their recognition memory will be affected by the spatial or gender property encoded in the L2.

3B. L2 learners will have a worse memory for the property encoded in their L2 than L1 speakers of that language.

4B. L2 proficiency will mediate the effect of the L2 on recognition memory. Lower proficient L2 learners will have worse recognition memory for the property encoded in their L2 than higher proficient L2 learners.

The main results can be summarized as follows. First we compared how often L1 German and L1 Spanish speakers recognized changes in object orientation and facial gender of agents. We found that L1 German speakers were significantly better to spot changes in object orientation (*M*=22% of the cases) than L1 Spanish speakers (*M*=6 %). We found that both L1 Spanish speakers (*M*=6%) and L1

German speakers (*M*=8%) hardly spotted any changes in the facial gender of agents (*M*=6%). Thus, we found partial support for Hypothesis 1B. Second, we compared L1 and L2 German speakers´ accuracy scores, with L2 learners divided over three proficiency groups (A2: beginner; B1=intermediate and B2+=advanced). We found that L2 German speakers (A2, *M*=38%; B1, *M*=37%, B2+, *M*=30%) were as good as L1 German speakers (*M*=22%) to recognize changes in object orientation. There were no significant accuracy differences between groups. Reaction time analyses and a comparison of accuracy on trials with spatial language and trials without spatial language further supported the idea this effect was caused by German placement verbs. There were no significant reaction time differences between groups, and trials with spatial language yielded significantly higher accuracy as compared with trials without spatial language. Thus these data supported Hypothesis 2B, but not Hypothesis 3B and 4B. We also compared L1 and L2 Spanish speakers´ accuracy scores. We found that both L1 Spanish speakers (*M*=6%) and L2 Spanish speakers (A2, *M*=12%; B1, *M*=12%, B2+, *M*=4%) hardly recognized changes in facial gender of agents. Differences between groups for the different picture changes were not significant. Thus we provided no support for Hypothesis 2B, 3B and 4B.

In the next chapter we elaborate on possible explanations for the positive findings for spatial language but negative findings for gender language and L2 proficiency. One explanation for the discrepancy between spatial and gender language may be that L2 German learners were instructed on spatial language before the experiment whereas L2 Spanish learners were not instructed on gender language. The reason for this was that we assumed L2 Spanish learners already knew the critical gender forms as they are presented in L2 Spanish teaching materials at beginner level. As all L2 German learners received the instruction, the instruction factor may also explain the lack of L2 proficiency effects.

In Paper 3 we considered whether L2 learners of German and Spanish rely on mental simulation as a route to comprehending their L2. We compared their data with that of L1 speakers of German and Spanish. In a computer experiment L1 and L2 speakers of German and Spanish read sentences that contained forms that vary cross linguistically across German and Spanish with respect to the expression of space and size of objects. They also saw pictures depicting objects in horizontal and vertical position and in prototypical and large size. We examined the following hypotheses:

1C. L1 and L2 German speakers will simulate object orientation whereas L1 and L2 Spanish speakers will simulate object size.

2C. Individual differences in L2 input and L2 learner factors will affect how fast L2 learners process meaning.

Also, we predicted that:

3C. L1 speakers will process meaning faster than L2 speakers.

The main results were as follows. First, we compared L1 and L2 German speakers´ responses to sentence-picture pairs where object orientation described in a sentence either matched or mismatched the position of a depicted object presented after reading the sentence. We found no significant differences in reaction times to match or mismatch trials. We did find that L1 German speakers RTs for both match and mismatch trials were faster than L2 German speakers´ RTs. We also compared L1 and L2 Spanish speakers´ responses to sentence-picture pairs where object size described in a sentence either mis/matched the size of a depicted object presented after reading the sentence. We found RTs for match trials were significantly faster than for mismatch trials. In addition, L1 Spanish speakers´ RTs for both match and mismatch trials were faster than L2 Spanish speakers´ RTs. Thus we found partial support for Hypothesis 1C. In particular, we found support that L1 and L2 Spanish speakers made mental simulations of object size. We also found support for Hypothesis 3C as German and Spanish L1 speakers showed faster RTs than L2 speakers. In the next chapter we elaborate on discrepancies in the results for object orientation and object size. They may relate to the salience of the different domains as such or experimental task demands.

We also investigated whether the following factors affected RTs in L2 learners: number of years spent learning the L2 (LeaYrs); pre- (Pre) and posttest (Post) scores on knowledge of target forms; number of hours of L2 use per day (DUH) and the average rate of perceived enjoyment, importance of and confidence in the L2 (AECI). For L2 German speakers we found effects for LeaYrs and Post and the interaction between Pre and Post. For L2 Spanish speakers we found effects for Post and the interaction between Pre and Post. Thus only some individual differences were found to affect how fast L2 learners processed meaning and Hypothesis 2C could only be partially supported by our data. In the next chapter we elaborate on why we found no effects for DUH and AECI. The lack of effects may be explained by lack of variation within the DUH and AECI data.

4.2 Summary of Findings per Cross Linguistic Difference

In this section I summarize the findings in this thesis for the three German-Spanish cross linguistic differences we investigated. Per linguistic domain I give an overview of the different cognitive functions that we investigated as well as findings for L1 speakers and L2 learners.

The predominant domain of investigation was (motion in) space. In particular, we investigated German placement verbs that indicate the position of a Figure object in relation to the Ground, in contrast with Spanish verbs that do not contain such spatial information. In Paper 1,2 and 3 we examined whether and how these verbs affected mental processes in L1 and L2 speakers. The cognitive functions that we investigated were categorization, recognition memory and mental simulation. As outlined in Table 2, we only found that German verbs affected L1 and L2 German speakers´ recognition memory, but no effects for categorization and mental simulation were documented.

Table 2: Summary of language-perception effects for the (motion in) space domain.

Space marked by German Placement Verbs			
Paper	**Cognitive Function**	**Effect L1 Speakers**	**Effect L2 Learners**
1	Categorizing	No	NA*
2	Memorizing	Yes	Yes
3	Simulating	No	No
Gender marked by Spanish Personal Pronouns			
1	Categorizing	No	NA*
2	Memorizing	No	No
Size marked by Spanish Augmentative Suffixes			
3	Simulating	Yes	Yes

* Not Applicable.

A second domain of investigation was gender. This domain was investigated in combination with Space in Paper 1 and 2. In particular, we considered Spanish personal pronouns that indicate gender of agents in comparison with German pronouns that do not indicate gender of agents. In Paper 1 and 2 we investigated whether these differences affected how L1 and L2 speakers categorized and memorized gender aspects. As outlined in Table 2 we found that the chosen Spanish-German cross linguistic differences in the expression of gender did not affect speakers´ categorization or memory of gender. Note that in Paper 2 we did find effects of spatial language on recognition memory. This discrepancy will be elaborated on in the following chapter.

The last domain we investigated was size, which was investigated in Paper 3 in combination with Space. In particular, we considered Spanish augmentative suffixes that indicate large size, in comparison with German, that cannot indicate large size by means of adding a suffix to nouns. We investigated how L1 speakers and L2 learners comprehended Spanish augmentative suffixes and whether these led them to make mental simulations of object size. As outlined in Table 2 we found support for the idea that these suffixes led both L1 and L2 speakers of Spanish to make mental simulations of size. Note that in Paper 3 we did not find that spatial language led to mental simulation of object orientation. This discrepancy will be elaborated on in the following chapter.

5. Discussion

The aim of this chapter is to discuss the empirical findings presented in Chapter 4 in light of the theoretical paradigms, challenges and questions sketched in Chapter 2. The goal of this thesis was to test three theories to make theoretical advance in the understanding of language-perception interaction in L1 and L2 speakers. In particular, we designed three studies to test the Sapir-Whorf and the Thinking for Speaking Hypothesis and notions of Grounded Cognition theory. In the following paragraphs I discuss to what degree the studies in this thesis have brought us closer to theoretical advance. In the final paragraph I give recommendations for future research.

5.1 Sapir-Whorf Hypothesis and Thinking-for-speaking

5.1.1 Theoretical Advances (Paper 1 and 2)

The first theoretical challenge that we defined related to the use of categorization tasks in Whorfian or TFS studies (Paper 1). We observed that studies typically employ categorizations tasks that investigate one conceptual domain (e.g. space). We also observed there are two types of categorization tasks that may reveal fine-grained (similarity judgment on a scale) and coarser differences (forced choice) in categorization. Finally, we observed that the categorization of (motion in) space and gender in placement events has not been investigated yet (but see Bosse & Papafragou, 2010 for a related investigation on space indicated by German positional verbs). Taking into account these observations, we first argued that strong support in favor or against Whorfian or TFS effects could be found if a study investigates multiple domains (e.g. space and gender) and language effects can be explained both ways (e.g. Language 1 shows an effect for space, but not for gender and Language 2 vice versa). Second, we reasoned that using two different types of categorization tasks, similarity judgment by scale and categorization by forced choice, to the same domains would provide us with both fine-grained and coarser categorization information.

In Paper 1 we designed a study that investigated cross linguistic differences and two types of categorization processes in relation to the conceptual domains (motion in) space and gender in the expression of placement events. This study yielded double support against Whorfian effects in two conceptual domains, by means of two different categorization tasks. Our null results are in line with results found in studies on motion verbs by Gennari et al. (2002), Papafragou, Massey and Gleitman (2002) and Cardini (2010). Importantly, we offered a plausible explanation for the null results by analyzing linguistic descriptions of stimuli by participants that were given after completion of the categorization tasks. Linguistic analyses confirmed the prediction of Papafragou, Massey and Gleitman (2002) that alternative descriptions may result in similar conveyance of semantics across languages despite cross linguistic differences. Divergent results by Papafragou and Selimis (2010) for motion verbs may be explained by the fact that their task implicitly encouraged language use. In their case,

participants may have actually employed the predicted linguistic forms. Divergent results by Athanasopoulos and Bylund (2013, 2014) may be explained by the following. Instead of manner of motion or different object positions resulting from motion their studies focused more on the domain of time (aspect) than to motion. Lera Boroditsky and colleagues (2002; 2003) found effects for categorization that were related to gender encoded in object nouns and not to gender descriptions of human agents as in this study. It may be that the linguistic instances under investigation in these studies have a more pervasive effect on cognition than those in our study. This could be explained by different saliency of linguistic forms as determined by factors as their frequency of occurrence, word class and their typical position in sentences as discussed in §2.4.4 (Talmy, 2000; Ellis, 2004, 2008). Let us take frequency of occurrence as example. Berman and Slobin (1994) argue that the less frequent forms appear in the input the less salient they are. Spanish is a "pronoun-dropping" language, which means pronouns may be omitted from speech or writing unless they are needed for purposes of emphasis or contrast (Butt & Benjamin, 1998). Thus it is plausible that the Spanish object nouns used in the studies of Borodistky and colleagues appear more frequently in language input than the personal pronouns *ellos* [they-masculine] and *ellas* [they-feminine]. The difference in salience of critical forms in the separate studies could explain the discrepancy in results.

The second theoretical challenge that we defined related to TFS language-memory studies with bilinguals, of which there are few (as pointed out by Filipovic, forthcoming). We observed there are no studies investigating placement events, combining two different domains (e.g. space and gender) and employing a bidirectional design (e.g. German learners of L2 Spanish and Spanish learners of L2 German).

In Paper 2 we set up such a study employing a fine-grained task as in Coventry et al. (2010) that could reveal effects of language on recognition memory. This study showed that the German language lead L1 and L2 German speakers to recognize changes in object orientation, whereas the Spanish language did not accommodate this advantage for L1 Spanish speakers and German learners of L2 Spanish. Interestingly, Spanish L2 learners of German' recognition memory was as good as that of L1 German speakers. For gender we found no such effects. The orientation results comply with previous findings on motion verbs (Billmann et al., 2000, Gennari et al., 2002); and motion verbs in L2 speakers (Filipovic, 2011) and can be interpreted as evidence in favor of TFS. The lack of gender effects does not comply with effects of language on memory for gender as reported by Boroditsky, Schmidt and Philipps (2003). However, this study did not investigate personal pronouns, but grammatical gender encoded in object nouns. As was the case in Paper 1, possibly, a higher saliency (determined by factors as frequency of appearance in input, word class, position within sentence etc.) of linguistic forms may have driven the effect in the study by Boroditsky, Schmidt and Philipps (2003) whereas lower saliency of linguistic forms in our study lead to no effects (see explanation above).

There are at least two (further) explanations for the lack of an effect for gender, but a positive result for spatial language. A plausible explanation concerning the L2 learners is that the instruction factor was not constant. This could arguably have resulted in a situation where the L2 German speakers' attention to the linguistic differences was greater than that of L2 Spanish speakers. Hence, we found an effect for L2 German and not for L2 Spanish. An alternative explanation is the following. There is both experimental and neurological evidence indicating that humans have a specific memory system dedicated to human faces that may qualitatively differ from other types of memory (Beyn & Knyazeva, 1962; Yin, 1969; Engstler & Engstler-Schooler, 1990). In fact, Engstler and Engstler-Schooler (1990) found that descriptions of faces worsened individuals´ memory for characteristics of those faces. Thus, a different nature of recognition memory for (the position of) objects and (gender characteristics of) faces may also explain the discrepancy between space and gender results. The findings of the present study do not allow choosing between the two proposed explanations, but future research could investigate to what degree they explain our findings.

5.1.2 Limitations

Considering Paper 1, the following limitations can be pointed out. It is plausible that performance on the similarity judgment task affected performance on the categorization by forced choice task. Also, given the repeated changes in object position and gender of agents in the categorization tasks it is perhaps not surprising that both German and Spanish participants employed alternative linguistic strategies from the ones predicted that marked object position and gender in the linguistic description task. Therefore, in future studies, one could consider finding different participant to complete the different categorization and linguistic tasks, although the downside of this method is inter-subject variation. Also, despite instructions, it may be that some participants perceived the placement stimuli as static instead of dynamic events as indicated by their use of static verbs, such as "hold", to describe the scenes, which may have affected their categorization responses. In future work on could consider the use of dynamic stimuli, although the downside of dynamic stimuli is that they are less controlled as compared with static stimuli.

Further, the investigation of two domains in Paper 1 brings along the issue of comparability. This applies to both linguistic instances as (visual) experimental manipulations. Considering the relevant linguistic instances (as described in Chapter 2, paragraph 2.4; and Paper 1) both verbs describing placement and personal pronouns describing human agents seemed comparable in the sense that they are basic and culturally significant productive lexical fields. Considering visual manipulations, the chosen cross linguistic differences involved manipulations that concerned both objects and human agents. For objects, we chose to change their position in space (horizontal vs. vertical) as in a related study by Bosse and Papafragou (2010). For human agents, we chose to replace the male agent by a female agent as in a relatively related study by Boroditsky, Webb and Philipps (2003). The clothing (both agents were wearing the same black shirt) and posture of both agents were kept as similar as

possible, yet aspects as body size, facial features, hair length and hair color differed between agents. The single orientation change of the object versus the multiple changes for the agent may have challenged the comparability of the different manipulations. In hindsight, a different manipulation applied to the same entity (the object) may have provided a better case for comparison. For example, in a future study one could manipulate the size of objects following the German-Spanish cross linguistic difference in the use of augmentative suffixes, as described in Chapter 2, paragraph 2.4.

In Paper 2 the issue of comparability of manipulations in picture stimuli is less profound as compared with Paper 1. In Paper 2, we employed the same cross linguistic differences as in Paper 1. Yet this time we employed more fine-grained visual manipulations as compared with Paper 1. First, we manipulated object orientation by changing the angle of an object held at an angle of 45 degrees, with 20 degrees up- or downward, following gradual changes employed by Feist and Gentner (2007) and Coventry et al. (2010). Second, we manipulated facial gender by morphing facial shape gradually. In a pilot study, we pinpointed the levels of the changes. These were piloted to be noticeable, yet not too easy to discriminate. Arguably the pilot testing reduced the chance that the facial changes were not comparable with the orientation changes in terms of difficulty (e.g. more difficult to distinguish than the orientation changes).

A more serious limitation in Paper 2 is that the instruction factor was not constant between L2 German and L2 Spanish learners. L2 German learners received a linguistic instruction that focused on German placement verbs, whereas L2 Spanish learners did not receive such an instruction in relation to Spanish personal pronouns. The reason for this was that the Spanish *ellos*/*ellas* [they-masculine/-feminine] are typically introduced to L2 learners in study material at A1-level (CEFR) whereas the German *legen/stellen* [lay/stand] are not. We came to this conclusion by talking with teachers and investigating the study materials *Instituto Cervantes* and *Goethe Institut* employ. The discrepancy in instruction arguably resulted in a situation where through "priming" L2 German speakers' attention to linguistic instances was greater than that of L2 Spanish speakers. Priming is an implicit memory effect in which exposure to one stimulus affects the response to another stimulus (Meyer & Schvaneveldt, 1971). In future studies, instruction conditions should be kept similar across different learner groups to maintain the largest degree of comparability.

5.2 Symbolic and Grounded Cognition

5.2.1 Theoretical Advances (Paper 3)

The first theoretical challenge that we defined was the following. We noted that research on mental simulation focused on object properties implied by a sentence context, yet properties are sometimes encoded in language itself. This led to question whether object orientation marked by German verbs leads to simulation match effects in German speakers. In addition, we asked whether information about large object size in Spanish augmentative suffixes would lead Spanish speakers to make simulations of

object size, as shown by match effects. The second theoretical challenge dealt with mental simulation in L2 learners. Do they rely on abstract symbols during language comprehension, or do they make mental simulations of object properties? A study by Tomczak and Ewert (2015) suggested that proficient L2 learners make mental simulations of motion in their L2.

In Paper 3, we set up a study that could address both theoretical challenges. We designed a study with German and Spanish sentences describing placement events. The German sentences contained the German verbs *legen* [lay] or *stellen* [stand] that marked vertical or horizontal object orientation. The Spanish sentences contained nouns indicating normal object size and nouns with augmentative suffixes that indicated large object size. These sentences were followed by pictures of objects in horizontal, vertical position and large and small size. We conducted the study with L2 learners of German and Spanish, as well as with German and Spanish monolingual controls.

Our first hypothesis posed that L1 and L2 German speakers would simulate object orientation whereas L1 and L2 Spanish speakers would simulate object size. We found no evidence (no match effect) for German orientation simulations in neither L1 nor L2 speakers, but we did find a match effect for Spanish size, in both L1 and L2 speakers, suggesting that both groups were making size simulations.

There are at least two accounts to explain the null effect for orientation and the match effect for size. First, both the visual and the linguistic stimuli may have differed in their salience across the two experiments. Considering the visual stimuli, Connell (2005, 2007) points out that psychophysically, properties such as shape or size, are more salient than orientation. Zwaan and Pecher (2012) also observe that orientation simulations are less robust than shape simulations. Considering the linguistic stimuli, Talmy (2000) has argued that semantic components expressed in main verbs are generally backgrounded and attract limited direct attention. This suggests that German verbs may have been less salient than Spanish suffixes. Thus, our results are in line with previous suggestions that propose that more salient properties such as size are being simulated whereas less salient properties´ simulation is less stable. Experiments by Rommers et al. (2013) did not support simulation of object orientation either (but see Zwaan, 2014).

Second, we may consider task demands. Remember that participants had to answer the question ´Was this object mentioned in the preceding sentence Yes or No?´. To complete the task effectively, participants only had to focus on the object nouns when reading the preceding sentences. If German participants applied such a strategy (which some participants reported in a debriefing) they might not have processed the orientation cue on the verb (i.e. *legen* [lay] or *stellen* [stand]). In contrast, it is likely that Spanish participants did process the size cue, since it was morphologically attached to the critical object nouns.

There is another explanation for the orientation null effect that relates to task demands. Remember that

our participants gave answers by pressing the Q(Yes) or P(No) buttons on a keyboard. Pressing buttons on a keyboard involves up-and downwards motion of fingers to a key (on a vertical axis) and coordination of left- and right finger pressing a left or right key (on a horizontal axis). The Theory of Event Coding (Hommel et al., 2001) predicts that when one motor response is completed (such as running a vertical orientation simulation needed to comprehend the sentence), the (vertical) feature will be bound to that outcome, making it temporarily unavailable (or less available) to other responses (such as the downward motion needed for the button press). Under such circumstances, priming between motor responses may be eliminated or reversed (Hommel et al., 2001). When priming is eliminated, a null effect may appear in a task as ours.

A third hypothesis posed that L1 speakers would process meaning faster than L2 speakers. Indeed, both for German and Spanish, L1 speakers showed faster RTs than L2 learners for both match and mismatch trials. This supports the idea that L2 learners may have weaker FMCs than L1 speakers (VanPatten, Williams and Rot, 2004). Slower L2 RTs might also be explained by involuntarily co-activation of L1 words during L2 sentence comprehension (Dijkstra, 2005).

5.2.2 Limitations

Considering Paper 3, the following limitations can be pointed out. First, the linguistic marking makes object orientation and size fully explicit and arguably relevant in completing the sentence-picture verification task. Remember the critical question in trials was ´Was this object mentioned in the preceding sentence Yes or No?´. If a sentence explicitly described a large lipstick, whereas a small lipstick is shown on the picture after, the picture invites a No response rather than a Yes response. This could induce response uncertainty leading to long response latencies. In this sense our study diverges from the simulation studies conducted by Zwaan and colleagues, where object properties (orientation, shape, color) were only implied by sentence context and irrelevant to complete the sentence-picture verification task. It is plausible that for some participants the linguistic marking induced No responses to mismatch trials. In fact, this was shown by considerable amounts of No responses to mismatch trials, which are reported in Paper 3. However, as we only analyzed Yes responses, and removed extreme outliers from the dataset, response uncertainties do not conflate the results in Paper 3. Therefore, we argue that our results are still comparable with previous simulation studies.

A second limitation is that data trimming procedures resulted in considerable loss of data. We followed trimming procedures designed by Stanfield and Zwaan (2001) and Zwaan and Pecher (2012). First, this involved the appliance of arbitrary cutoffs, which meant that original means and standard deviations were replaced by faster means with lower variance. Second, a focused trimming procedure was applied, where means and standard deviations were calculated for each participant. RTs further away than two standard deviations from individuals´ means were trimmed. Finally, we analyzed median RTs. These substantial procedures decrease how reflective results are as compared with the natural data. Yet,

comparability with the many previous simulation studies was of crucial importance to us and outweighed the limitations the trimming procedures brought along. Thus, after consideration, we chose to keep trimming procedures comparable with previous studies.

5.3 L2 Learner Differences in Cognition

5.3.1 Theoretical Advances (Paper 2 and 3)

The first theoretical challenge we defined related to the question whether visual memory after verbal-visual encoding in L2 learners is affected by L2 proficiency. Remember we discussed that research has shown that L2 learners with higher levels of L2 proficiency process sentences almost as automatic as L1 speakers whereas speakers at lower levels of L2 proficiency do not (Rossi, Gugler, Friederici & Hahne, 2006; Van Hell & Tokowicz, 2010; Foucart & Frenck-Mestre, 2013). This suggests that L2 processing imposes a higher cognitive load on L2 learners with lower L2 proficiency as compared with L2 learners with higher L2 proficiency and L1 speakers. It has been shown that a higher cognitive load caused by performing multiple tasks simultaneously impairs both recall and recognition (e.g. Hicks & Marsh, 2000; Naveh-Benjamin, Craik, Guez & Dori, 1998). Considering these findings on L1 and L2 processing, cognitive load and memory impairment, we predicted that L2 learners with lower L2 proficiency will have a worse recognition memory than L2 learners with higher L2 proficiency. In Paper 2, we tested L2 learners of German and Spanish divided over three groups of L2 proficiency (beginner, intermediate, advanced) on the basis of their self-reported levels in terms of the CEFR. We compared memory performance of these groups.

In Paper 2 we found that lower and higher proficient L2 learners´ recognition memory for object orientation was as good as that of L1 speakers. This was not in line with predictions about higher cognitive load in L2 speakers in general and at lower levels of L2 proficiency in specific and consequent effects on memory. Findings suggest that in this experiment, the cognitive load of language processing was equal for L1 speakers and lower and higher proficient L2 learners, leading to similar recognition memory accuracy. How can we reconcile our findings with research that indicates L2 processing differs at different levels of proficiency, with higher proficient learners approaching L1 speaker automaticity (Rossi, Gugler, Friederici & Hahne, 2006; Van Hell & Tokowicz, 2010; Foucart & Frenck-Mestre, 2013)? We may explain the discrepancy as follows. It is important to realize that in research on L2 processing participants are usually presented with sentences that contain syntactic or semantic anomalies and differences in electrical brain activity to those anomalies have been shown. In our experiment, participants were presented with sentences that were grammatically correct and semantically unambiguous. Moreover, all learners received the L2 instruction on critical German verbs right before embarking on the memory experiment. This may have resulted in a situation where possible differences in knowledge of these verbs and their semantics before the experiment were minimized. Therefore it may be argued that sentences and critical verbs were more or less equally easy to process for both lower and higher proficient L2 learners. If this was the case, it is not surprising that

we did not find the subsequent differences for recognition memory that we predicted for lower and higher proficient L2 learners and L1 speakers.

The second theoretical challenge we defined related to the question whether individual differences in L2 learners affect how they process meaning and make mental simulations. In particular, we investigated whether differences in L2 learners´ proficiency, prior knowledge about the target forms, L2 use and motivation affected their speed of reaction. The literature on language processing suggests that in early stages of L2 learning, lexical items are processed through association with their translation equivalents in the L1, whereas in later stages, with higher proficiency, processing of L2 words is more directly conceptually mediated (Dufour & Kroll, 1995; Kroll & Stewart, 1994). Thus, we predicted that the higher the L2 proficiency, prior knowledge of target forms, L2 exposure and motivation, the faster L2 learners would process meaning and the bigger the chance that they would rely on mental simulation as a route to comprehension, instead of relying on abstract symbols to extract meaning. In Paper 3, we tested L2 learners of German and Spanish, and recorded information on the different background factors we were interested in. We constructed a multilevel regression model. Exploratory analyses revealed no effects of CEFR L2 proficiency. Instead, we considered years of L2 instruction as indicator of general L2 proficiency. Also we measured knowledge of the target forms by a pre- and posttest before and after L2 instruction. Finally we considered L2 use in hours per day and we computed a measure of motivation by collapsing ratings to three motivation statements.

In Paper 3 we found an effect for general L2 proficiency for L2 German learners, where higher proficient learners reacted faster to stimuli than lower proficient learners. No such effect was found for L2 Spanish learners. This may be due to the larger amount of data points included that we had for L2 German as compared with L2 Spanish. The smaller amount of data points for L2 Spanish may have led to less statistical power for detecting an effect. As for L2 knowledge about target forms, we found both for L2 German and L2 Spanish learners that students with lower scores in the pretest (indicating no or little knowledge before the L2 instruction) and posttest (indicating they did not understand the L2 instruction properly) responded the slowest to critical stimuli. Arguably, students with higher pretest scores had prior knowledge of the target language forms before participating in the study. Potentially, this knowledge was consolidated by instruction, thereby leading to more solid form meaning connections (VanPatten, Williams & Rott, 2004). Thus, a fine-grained measure as reaction time (as opposed to a measure as recognition memory accuracy in Paper 1) was able to reveal faster reaction times for students with higher pretest scores as compared with students with lower pretest scores.

We found no significant effects of L2 use and motivation on performance. This indicates that such effects may be non-existent. However, the nature of our datasets may also have caused the lack of effects. First, we observe that mainly high motivated L2 learners participated in our experiments. The smaller number of learners at the lower end of the scale may have taken away statistical power to reveal significant differences. Second, we observed that the largest number of our learners did not use their L2 for many hours per day. This is not surprising given that we tested students in a foreign

language context. These L2 students may have difficulties in finding L1 speakers in their environment to interact with on a daily basis. In the questionnaire (Gullberg & Indefrey, 2003) L2 German speakers reported to speak their L2 0,5 hours per day and L2 Spanish speakers 0,6 hours per day. They reported to speak the L2 with relatives, partners, friends, colleagues, classmates, roommates, but did not report whether these conversation partners were L1 speakers or not. Further, both in Germany and Spain, movies and TV-series are dubbed in German and Spanish respectively (Whitman-Linsen, 1992), which takes away opportunities for L2 students to hear their L2 outside of class. It is plausible that the lack of variance in the data we collected in a foreign language context decreased the chance to detect an effect for L2 Use.

5.3.2 Limitations

Considering Paper 2, we find the following limitation: the instruction factor was constant for L2 German learners at different levels of proficiency. The reason for this was that we wanted to ensure that all participants knew the critical forms before completing the memory task. However, instructing all learners right before the memory task plausibly caused all learners to have active form-meaning connections that ensured easy processing. As the simple sentences where the critical forms occurred did not require much effort to process either, the chance to find differences for learners with different levels of L2 proficiency may have been eliminated. In future studies one may consider presenting learners with extended text, which may increase processing demands for L2 learner with lower levels of L2 proficiency. Also, one may consider differentiating between learners who already know the form and those who don't and only instruct the latter group. In this case proficiency differences may be revealed for learners who have just learned the critical forms in comparison with learners that already knew the forms for longer periods of time.

Considering Paper 3, we found that the German and Spanish L2 learner groups were not fully comparable. They differed as to how many learners with different L2 proficiency they contained. Moreover, after applying data trimming procedures, we had more data points that we could analyze for the L2 German learners as compared with the L2 Spanish learners. This may have increased power to find an effect for L2 proficiency for L2 German, but not for L2 Spanish. Also, we found that there was limited variance in L2 exposure for both German and Spanish learners and mainly highly motivated learners participated in this study. The limited variance for L2 exposure and motivation may have decreased the statistical power to find effects for these factors.

5.4 Directions for Future Research

Directions for future research can be made from at least three perspectives. First, recommendations can be made for further work in relation to the three theoretical paradigms that have been discussed in the thesis. Second, we can discuss which of the conceptual domains, (motion in) space, gender and size,

under investigation in this thesis is most promising in terms of uncovering language-perception interactions. Finally, we can make recommendations for the study of language and perception in theoretical paradigms and with methods that differ, yet complement those discussed in this thesis. We discuss each of these three perspectives in more detail below.

We can make recommendations for further work within the three theoretical paradigms that have been discussed in the thesis. With respect to the Whorfian and TFS research paradigms we addressed, we make the following recommendations. We argue that Whorfian categorization tasks without explicit offering or production of language are not as informative as desired since it is not possible to determine if participants employ critical language as a strategy to complete tasks (Papafragou, Massey, Gleitman, 2002). In future studies, tasks with overt receptive or productive use of language (as in a TFS paradigm) may be more informative as to whether L1 speakers of different languages categorize according to cross linguistic differences in their languages. In addition, they could inform us as to whether L2 speakers of such languages categorize based on distinctions made in their L1, their L2 or whether they adhere to another pattern (see Athanasopoulos et al., 2015 as a recent example of such a study). We also encourage more TFS work on language and memory with bilinguals, as there is not abundant work available (as remarked by Filipovic, forthcoming). The study presented in this thesis (Paper 2) shows that in the domain of (motion in) space, effects of language on recognition memory both in L1 and L2 speakers may be predicted. Further empirical work needs to unravel and consolidate whether these findings only apply to spatial language or extend to other conceptual domains encoded within language.

Robinson and Ellis (2008) point out that most studies dealing with TFS and L2 learning lack a clear integration with models of language representation, processing and production (see Kroll & Tokowicz, 2005 for an overview). For example, it has been proposed that TFS may have important implications for models of speech processing, such as Levelt´s (1989) blueprint of the speaker (Carroll, 1994; Glatz & Von Stutterheim, 2003; Treffers-Daller & Tidball, 2015). TFS implies that language affects thoughts processes at the conceptualization stage, which means that Levelt´s model needs to be revised in view of the evidence from studies that support cognitive restructuring in L2 learners and bilinguals. We propose it may be interesting to integrate TFS and the Revised Hierarchical Model (Kroll & Stewart, 1994; Kroll et al., 2010) of bilingual representation and extend the model with a recognition memory component. Let us take cross linguistic differences in the expression of object position in German and Spanish and Spanish learners of L2 German, which were investigated in Paper 2, to illustrate this idea. For Spanish learners of L2 German, in an extended RHM model, the Spanish verb *poner* [put] would have a weak or no link to recognition memory for the position of objects in space. However, the L2 German verbs *legen* [lay] and *stellen* [stand] would have a link of considerable strength to recognition memory for object position. The strength of these L2 form-recognition memory links may vary according to level of L2 proficiency, yet such differences could not be revealed in Paper 2.

In relation to grounded cognition in general and mental simulation in particular we encourage further exploration of several research lines. Two lines that conceptually extend work on mental simulation have been touched upon in the work presented in this thesis. First we have addressed that object properties can be implied by a situation sketched in a sentence or they can be marked by linguistic forms (verbs, suffixes). At least, we have shown that Spanish augmentative suffixes may lead to simulation of object size. An important follow up question could address whether different types of information (implicit vs. explicit) lead to simulations that are less or more complete, leading to slower or faster reaction times in an experimental task as the one we employed. For example, one could design a study where participants read sentences that imply object size ("From the airplane in the sky Harry looked at the house", implied size of the house: small) and sentences that make object size explicit by means of augmentative or diminutive suffixes and determine whether both type of sentences lead to match effects and if so whether RTs differ.

Another line to which we have contributed in this thesis investigates whether L2 comprehension involves simulation (Vukovic & Williams, 2014; Tomczak & Ewert, 2015). Taking a pluralist view of cognition, as proposed by Zwaan (2014), may reconcile opponents of symbolic versus grounded cognition. A pluralist view holds that both abstract and grounded symbols to language comprehension may vary as a function of the degree to which language use is embedded in the environment. The more that language use is embedded in the environment, the larger the chance that grounded symbols are used as a route to comprehension. In Paper 3 in this dissertation we examined whether L2 learners in an environment where the L2 is not the dominant language rely on abstract symbols or grounded, mental simulations during L2 comprehension. Moreover, we have addressed whether individual differences in comprehenders affect simulation (Vukovic & Williams, 2015). We found that knowledge of specific linguistic forms may lead to RT differences in L2 learners. A possible follow up study could consolidate this finding by comparing two groups of L2 learners: a group that receives and a group that does not receive instruction on particular linguistic forms. In case the L2 learners that receive instruction show a match effect whereas the L2 learners that do not receive instruction do not, this would provide further support for the notion that L2 instruction may facilitate learning and lead to mental simulation in L2 learners.

Other promising lines of research on mental simulation that have not been addressed in this thesis are the following. A promising line of research has started to look at whether mental simulations are being updated during sentence comprehension (Sato et al., 2013; Hoeben Mannaert, 2016). Also, researchers have started to explore whether L1 speakers simulate in extended discourse (Ditman et al., 2010). Replication studies that replicate simulation effects are also of crucial importance (Zwaan & Pecher, 2014; Rommers et al., 2013). We encourage each of the lines of research discussed here and above to be further developed in order to further understand mental simulation. In particular we encourage further research with L2 learners, as there is little research available currently.

We recommend caution when trying to uncover individual differences in L2 learners´ sentence comprehension and possible subsequent effects on other cognitive processes, such as memorizing. Our trials within this thesis were partly unsuccessful (see Paper 2 and 3). In future memory studies (Paper 2) on may consider to present learners with extended text, which may increase processing demands for L2 learner with lower levels of L2 proficiency. Also, one may consider differentiating between learners who already know the form and those who don't and only instruct the latter group. In this case proficiency differences may be revealed for learners who have just learned the critical forms in comparison with learners that already have known the forms for longer periods of time. In future simulation studies (Paper 3) one may aim once more to constitute balanced data sets that are able to reveal differences. For example, one could compare L2 learners in a foreign language context (with little L2 exposure) with L2 learners in an immersion context (with much L2 exposure). Another challenge in future studies, especially bidirectional studies, is to find proficiency tests that take specific measures of phenomena of interest and that are comparable for speakers of different languages.

Moving from the theoretical paradigms to the event type and conceptual domains of interest in this thesis, we recommend the following. We argue that placement events remain an interesting topic, as has been shown by recent interest of researchers from different disciplines (e.g. Cadierno et al., 2016; Gullberg, 2009; Van Bergen & Flecken, 2017). Therefore, we encourage this type of event to be investigated further. The conceptual domains we investigated in relation to placement events were (motion in) space, gender and size. (Motion in) Space remains an interesting domain to investigate, as has been shown by results in Paper 2 of this thesis. In Paper 3, we revealed that size, a relatively unexplored domain, may be a promising domain to explore further as we found promising results here as well. For gender we found null results both in Paper 1 and in Paper 2. As there is little research on the relation between gender in language (as pointed out by Pavlenko, 2014) and L1 and L2 speaker's nonlinguistic cognition, it is too early to exclude this domain from investigation. Rather we would encourage further refinement and investigation, especially considering the relevance of gendered language in the public debate (Szesny et al., 2016). Corpus analyses (as in Lemmens, 2006) focusing on frequency of occurrence of particular linguistic instances may proof helpful for researchers to select those linguistic instances that are encountered and produced frequently by speakers of different languages for investigations of the relation between language and perception.

Finally, we recommend (space, gender and size in) placement events to be further investigated with different methods (see Athanasopoulos & Bylund, 2013). In recent times, and as exemplified by the Marie Curie Network on Language and Perception (2013-2016), collaboration between researchers within linguistics, psychology and neuroscience has intensified. It is recommended to study topics that have been investigated behaviorally in neurological research designs, employing techniques such as electroencephalography (EEG) or functional magnetic resonance imaging (fMRI) to explore how language perception interaction is represented neurologically. In relation to placement events, Van Bergen and Flecken (2017) have studied placement events employing yet another technique, eye

tracking, investigating speakers of various world languages. Eye tracking reveals patterns of attention as it measures eye positions and eye movement when inspecting written language and pictures. Van Bergen and Flecken found that speakers of languages that mark object position through placement verbs (Dutch, German) start looking at objects that match the described object position as soon as they hear a placement verb. This was not the case for speakers of languages whose placement verbs do not indicate object position (French, English).

To conclude, a central challenge in all fields of cognitive science is to move the study of language from the micro (words and sentences) to the macro level (extended written texts and conversation) in an empirically valid way. Most of the work in psycholinguistics in general focuses almost exclusively on individual words and sentences (Fischer & Zwaan, 2008). Psychological experiments on language typically use decontextualized words and sentences as stimuli. Zwaan (2014) points out that this lack of context in language research is problematic because the resulting patterns of brain and behavioral responses obtained cannot be extrapolated to discourse comprehension. He claims that during discourse comprehension, people generate extensive mental representations, elements of which may form the context of an utterance and can override word-level patterns of association (Nieuwland & Van Berkum, 2006). He thus argues that only through the study of context we will make significant progress toward understanding language comprehension and cognition in general. He proposes to discern different forms of language comprehension in terms of how deeply they are embedded in the environment. Detailed analyses of levels of embeddedness are likely to yield novel predictions about the relative contributions of processes and representations in language comprehension. This is one valid proposal on how to move forward in terms of meeting the central challenge within cognitive science, that is, to explain how humans can understand language.

Paper 1:

Dietha Koster

Categorizing motion in space and gender in placement events – A cross linguistic comparison of German and Spanish speakers

Categorizing Motion in Space and Gender in Placement Events – A Cross linguistic Comparison of German and Spanish speakers

Dietha Koster

Abstract: Humans place objects onto surfaces so frequently that we hardly ever stop to think about it. However, different languages display variation in the way they describe the position of objects in these placement events (Kopecka & Narasimhan, 2012). For example, German employs the verbs *legen* [lay] and *stellen* [stand] that indicate horizontal and vertical object position whereas the Spanish verb *poner* [put] does not. German and Spanish also differ as to how they describe the gender of multiple human agents through personal pronouns (Hellinger & Bußmann, 2001, 2002, 2003). Spanish distinguishes between *ellos* [they-masculine] and *ellas* [they-feminine] whereas the German *sie* [they] does not mark gender. Can these cross linguistic differences affect the way speakers of German and Spanish categorize space and gender in placement events? Following a strong interpretation of the Sapir-Whorf hypothesis (Whorf, 1956) the answer would be yes, even in a task where individuals do not overtly produce or encounter language. In such a context individuals may still use language as a mental strategy (Gentner & Goldin-Meadow, 2003). However, Papafragou, Massey and Gleitman (2002) point out that alternative descriptions may result in similar conveyance of semantics across languages despite cross linguistic differences. In this paper, we examine these predictions by employing two types of categorization tasks and a linguistic description task. German and Spanish participants were given a series of pictures of placement scenes where object position and gender of agents varied. We asked participants to categorize the pictures by giving similarity judgments on a scale from 1-7 and by choosing two out of three pictures that were most similar. After we asked participants to provide descriptions of placement scenes. Results for both categorization tasks show no Whorfian effects for the categorization of object position or gender of human agents. The German and Spanish descriptions revealed that participants may have employed linguistic strategies during categorization different from the ones predicted. The descriptions marked object position and gender similarly across languages. These results indicate that cross linguistic differences do not affect how German and Spanish speakers categorize spatial and gender aspects in placement events. Moreover, they call for a cautious approach when theorizing about Whorfian effects in contexts where no overt language is present.

Key words: Sapir-Whorf hypothesis, Categorization, Placement Events, Motion, Space, Gender

Highlights:

- German-Spanish cross linguistic differences in the expression of placement events do not affect categorization of space or gender on two types of categorization tasks

- Despite cross linguistic differences, German and Spanish speakers mark space and gender equally often through alternative descriptions

1. Introduction

A classic debate within cognitive science is whether human thought is shaped by language (Gardner, 1985). The Sapir-Whorf hypothesis, whose origins can be traced back to Benjamin Whorf (1956), poses that (a) languages vary in their semantic partitioning of the world; (b) the structure of one´s language influences the manner in which one perceives the world; (c) therefore, speakers of different languages will perceive the world differently. The hypothesis has been interpreted in weaker ("thinking for speaking", Slobin, 1996) and stronger manners (see Wolff & Holmes, 2011 for a review). Following a strong interpretation of the Whorfian hypothesis, language is assumed to affect perception even in contexts where language is not overtly offered or produced. Gentner and Goldin-Meadow (2003) point out that even in context without language production or reception, humans may still use language as a mental strategy to complete a given task. The literature is unresolved about if and in which contexts the Sapir-Whorf hypothesis holds (Trueswell & Papafragou, 2010; Fausey & Boroditsky, 2011; Bylund and Athanasopoulos, 2014). Whether language may affect perception depends on a number of factors such as the specific characteristics of the domain involved, the nature of the particular linguistic feature under investigation and the degree to which the experimental task promotes or inhibits strategic use of linguistic categories (Bylund & Athanasopoulos, 2014).

Whorfian questions have been investigated empirically by addressing typological differences within various language domains. Cross linguistic differences in the expression of color, time, number, space and motion have received considerable attention within the Whorfian literature (for recent reviews, see Evans, 2011; Malt & Majid, 2013). A well-investigated cognitive process is that of categorization. This fundamental aspect of human cognition is considered to operate on the basis of similarity, so that two stimuli that are perceived as similar are likely to be classified as members of the same category (Nosofsky, 1986). A number of studies within the motion domain have provided evidence in favor of Whorfian effects in categorization without participants giving verbal descriptions of motion events. For example, Papafragou and Selimis (2010) showed that linguistic variation in the expression of manner of motion through verbs (e.g. skipping, jumping, crawling etc.) affected how speakers categorize motion events. Athanasopoulos and Bylund (2013) and Bylund and Athanasopoulos (2015) also found effects of cross linguistic differences on categorization preferences in a context where no language was produced. They found that linguistic variation in the expression of the temporal dimension of orientation towards goals affected how speakers categorized motion events.

Motion events that have received less attention are so-called "placement events". These are events where agents move an object to a certain location, as in: *They put their glasses on the table*. Speakers of different languages vary in the information they include in verbs when describing such events (Kopecka & Narasimhan, 2012). German and Spanish placement verbs differ as to whether they encode

the position of the object being placed. In German, one expresses whether the object ends up in a vertical (*legen*, [lay]) or horizontal (*stellen*, [stand]) position (Fagan, 1991; Lemmens, 2006). In Spanish, object orientation is not expressed through placement verbs (*poner* [put]; *dejar* [leave in a place]) (Ibarretxe-Antuñano, 2012; Cadierno et al., 2016). German and Spanish also differ in whether they give information about the gender of multiple agents, when describing them with personal pronouns. In Spanish, it is obligatory to mark whether a group of two or more agents consists of females only *ellas* [they-feminine] or whether it is a mixed or male-only group, *ellos* [they-masculine], whereas German offers a single pronoun *sie* [they] to describe either possibility (Cartagena & Gauger, 1989). This variation constitutes an interesting case with which to study the interaction between the linguistic description of actions and people and nonlinguistic cognition.

Papafragou, Massey and Gleitman (2002) have pointed out that alternative descriptions may result in similar conveyance of semantics across languages despite the existence of linguistic differences in verbs or pronouns between languages. "In the end one can translate the semantics of such words into other languages through phraseology that, though it may lose some of the color and natural force of the original, yet conveys the semantics passing well." (Papafragou, Massey & Gleitman, 2002: 193). For example, if a Spanish speaker would feel the need to specify the orientation of objects in a placement scene he could utter a sentence such as *Ellos ponen los vasos echados sobre la mesa* [They put the glasses lying on the table]. If a German speaker would want to express the gender of agents in a placement scene, he could produce an utterance such as *Der Mann und die Frau stellen die Gläser auf den Tisch* [The man and the woman stand the glasses on the table]. If speakers indeed employ such strategies when performing a categorization task, hypothesized effects of cross linguistic differences on categorization may be ruled out.

The main goal of this paper is to examine how German and Spanish speakers categorize the position of objects and gender of human agents in placement events. It is important to mention that unlike many previous studies, we investigate two domains, motion in space and gender, within one study, which may provide double support in favor or against the Whorfian hypothesis. Our focus is on how German and Spanish speakers categorize placement events, when language is not explicitly offered or used when performing two categorization tasks. Thus, we are probing a strict version of the Whorfian hypothesis, without including unnatural conditions (such as repeating nonsense syllables) that prohibit the possible use of language as a strategy during task performance. Also, we examine how German and Spanish speakers describe pictures of placement events where the position of objects and the gender of agents vary to determine whether our linguistic predictions were correct or not. In the following, we provide further theoretical background that will lead to the formulation of testable hypotheses.

2. Theoretical Background

2.1 Motion and Categorization – Methodological Considerations

To give an exhaustive overview of Whorfian studies across domains is not our aim. In this section we

discuss methodological considerations of work of immediate relevance to us, which are categorization studies in the domain of motion. Motion events have received considerable interest cross linguistically following Talmy´s typology of motion events (Talmy, 1975, 1985, 2000). Most categorization studies have employed tasks that involve categorization by forced choice. Here, triads of pictures (or videos) are presented and subjects are asked two select two pictures that are most similar (as in Gennari et al., 2002). Fewer studies have employed a task that involves similarity judgment of sets of two pictures on a scale (as in Boroditsky, Ham & Rascar, 2002). Arguably, similarity judgment on a scale gives a more fine-grained measure of cognitive preferences than categorization by forced choice.

An unresolved issue in the field is whether language preferences affect categorization in tasks that do not involve explicit use of language. With such "free encoding" tasks, participants are neither presented with language nor prohibited from language use when doing a categorization task. The danger with these tasks lies in the fact that one cannot be sure whether participants use language as a strategy during tasks or not, unless they are simultaneously engaged with unnatural tasks that prohibit the use of language, such as counting backwards or repeating nonsense syllables (as in Gennari et al., 2002). Conflicting evidence comes from studies on verbs that encode manner of motion and verbs that encode temporal descriptions of goal-oriented motion. Considering manner of motion, Gennari et al., (2002) and Cardini (2010) found no effect of cross linguistic differences with a "free encoding" categorization task. In both studies participants were asked to categorize motion events without being presented with or urged to produce language that described stimuli. Papafragou, Massey and Gleitman (2002) found no effect either, even though participants described stimuli a day before performing the categorization task. However, Papafragou and Selimis (2010) did find a language-specific effect for categorization of manner of motion with free encoding. More evidence in favor of the Whorfian hypothesis comes from work on typological differences in temporal descriptions of goal-oriented motion. Athanasopoulos and Bylund (2013), Bylund and Athanasopoulos (2014) and Athanasopoulos et al. (2015) all found language-specific effects for categorization with free encoding tasks.

Some factors have been put forward to specify the exact conditions in which effects are observed and when they do not appear. For example, it has been argued that language effects on categorization are more likely to appear when stimuli are complex (Filipovic, forthcoming). It is suggested that the integration of complex information presented in stimuli causes subjects to rely on language-specific lexicalization resources that would be activated as an aid (Fausey & Boroditsky, 2011; Trueswell & Papafragou, 2010). However, one may put into question how closely extremely complex stimuli and tasks reflect categorization processes that individuals experience in everyday life.

It is important to mention that none of the studies discussed above have been designed to investigate two conceptual domains simultaneously. Yet, arguably, a strong experimental design ensures that cognitive preferences caused by language can be explained both ways (e.g. Language 1 shows an effect for conceptual domain 1, whereas Language 2 does not; and vice versa for conceptual domain 2). A study by Boroditsky, Ham and Rascar (2002) is an example of a similarity judgment study where there

were two types of conceptual changes, yet only one of the changes could be explained by cross linguistic differences between English and Indonesian. In particular, temporal aspect in pictures (a man that is about to kick/is kicking/has kicked a ball) as well as human agents (man kicks a ball, different man kicks a ball) was changed in this study. They asked participants to rate similarity of different picture pairs (e.g. same tense, different actor and different tense, same actor). Language differences in the expression of temporal aspect in English and Indonesian explained differences in similarity ratings for the temporal changes. However, there was no language prediction involved concerning differences in similarity judgment of agent changes. In this paper we aim to design a study that examines cognitive preferences by investigating two conceptual domains, space and gender, simultaneously.

2.2 Space and Gender in Descriptions of Placement Events

Verbs of "putting" and "taking" are amongst the most frequent, basic verbs in a language and they are amongst the earliest verbs learned by children (Levinson, 2012). There are a few studies on typological differences in placement verbs and nonlinguistic cognition (eye movement, memory). Bergen and Flecken (2017) found that speakers of languages that mark object position through placement verbs (Dutch, German) start looking at objects that match the described object position as soon as they hear a placement verb. This was not the case for speakers of languages whose placement verbs do not indicate object position (French, English). Bosse and Papafragou (2010) investigated the effects of the positional counterparts of the German placement verbs *legen* [stand] and *stellen* [lay], which are *liegen* and *stehen*, on memory. *Liegen* and *stehen* also indicate a vertical versus horizontal position of an object, yet, they do not indicate that these positions result from motion. Results showed that *liegen* and *stehen* did not affect memory for object position in German speakers, even when they were urged to produce the critical verbs. Considering categorization of placement, Hae in Park and Ziegler (2015) found that linguistic differences in the expressions of "put in" and "put on" affected how speakers of English and Korean categorized placement events. More specifically, English and Korean speakers based their categorization preferences on the categories described in their native language, without verbalizing those.

Placement events often refer to the agents who are performing the placement action(s). In any language, personal nouns and pronouns are employed to communicate about the self and others and to identify people as members of male and female groups (Hellinger & Bußmann, 2002). There is limited experimental research on how speakers perceive gender expressed through personal pronouns. For example, Lamers et al. (2008) showed that event-related brain potentials (ERPs) of Dutch and German speakers differed when processing male, female and neutral pronouns. They related these differences to language specific characteristics concerning the Dutch neutral pronoun *het* [it]. In the Whorfian research tradition, some researchers have studied grammatical gender expressed in object nouns and categorization. For example, Philipps and Boroditsky (2003) found that grammatical gender encoded in object nouns affected German and Spanish speakers´ categorization of objects, men and women. This effect remained even when participants performed a verbal interference task.

Considering the theoretical review so far we may draw up the following observations. First, results of Whorfian studies in the domain of motion are inconclusive as to whether language affects categorization in a context with no overt language use. To our knowledge, there are no studies that examine two conceptual domains within a single investigation. Second, both placement verbs and personal pronouns constitute basic, culturally significant productive lexical fields. Empirical research has shown that placement verbs may affect nonlinguistic cognition in speakers of different languages and there is limited evidence that personal pronouns affect speakers´ nonlinguistic cognition. Third, to our knowledge, the effect of placement verbs or personal pronouns on categorization preferences has not yet been investigated. These three observations provide motivation for the present investigation. In particular, our main goal is to investigate whether typological differences in German and Spanish in the description of placement events (as described in the Introduction) cause differences in the categorization of object position and human agents in placement events. We investigate this by collecting both similarity judgments on a scale and categorization preferences by forced choice. We also examine linguistic descriptions of placement events. These descriptions may indicate whether participants employ the predicted descriptions or alternative descriptions. In the latter case, we investigate whether the alternative descriptions mark object position and gender.

2.3 Hypotheses

In the following we report on a study where we presented German and Spanish speakers with placement scenes that varied both in the position of objects as well as in the gender of agents. We asked these speakers to a) judge the similarity of two pictures depicting different placement scenes on a scale and b) categorize placement scenes by selecting two out of the three pictures depicting different placement scenes that are most similar. Based on the theoretical review above, we made the following predictions:

1. German speakers will give placement scenes with same object position higher similarity ratings than Spanish speakers. Spanish speakers will give scenes where the gender of agents is the same higher similarity ratings than German speakers.

2. German speakers will categorize placement scenes on the basis of similarity in object position; whereas Spanish speakers will categorize them on the basis of similarity of the gender of agents.

Finally we examined whether our expectations about cross linguistic differences in the descriptions of placement scenes by German and Spanish speakers were correct. These descriptions were given after the categorization tasks to exclude possible effects of linguistic encoding before doing the categorization tasks. In particular, we investigated whether German speakers described placement scenes where objects are placed in horizontal and vertical positions with the verbs *legen* [lay] and

stellen [stand] and Spanish speakers with the verbs *poner* [put] or *dejar* [leave in a place]. Also we examined whether Spanish speakers described female-only or mixed groups of agents in placement events with the pronouns *ellos* [they-masculine] and *ellas* [they-feminine] respectively and whether German speakers described these agents with the pronoun *sie* [they].

3. Method

3.1 Participants

22 German (9 male) and 23 Spanish (7 male) native speakers were paid a nominal fee to participate in the study. The German participants were students recruited at the University of Bremen whereas the Spanish speakers were students recruited at the University of Seville. Most participants reported knowledge of English and limited knowledge of other languages, but we ensured that German students had no knowledge of Spanish and Spanish students had no knowledge of German.

3.2 Materials

A set of 13 color pictures served as stimuli in the two categorization tasks in this study. The experimental pictures portrayed two people putting a cylindrical shaped object on a table or a koala (see Figure 1 and 2 for examples). The different objects were: lipstick, deodorant, tube of toothpaste, binoculars. These objects were normed in a pilot study to be orientation-free objects, which can naturally occur in a horizontal and vertical position[1]. For each experimental scene we manipulated the gender of one actor (male vs. female) or the position of the placed object (vertical vs. horizontal) as in a related studies by Boroditsky, Webb and Philipps (2003) and Bosse and Papafragou (2010). Thus we had four placement scenes plus 4 x 2 alternates for each scene is 12 pictures, plus one control koala picture. See Figure 1 for examples of three critical stimuli.

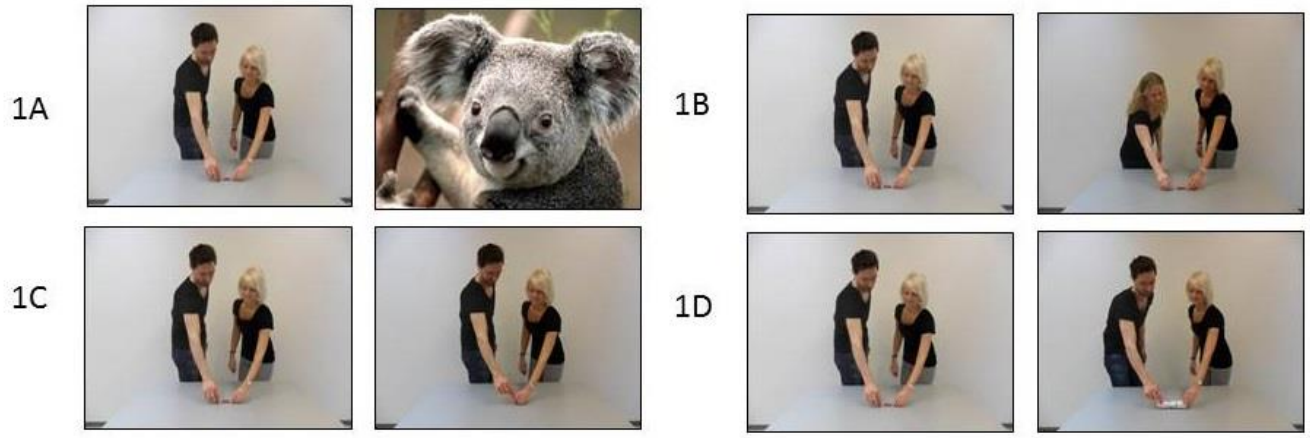
3.3 Design

Categorization by Similarity Judgment

A set of 36 experimental items was created: 4 placement scenes (lipstick, deodorant, tube of toothpaste, binoculars) x 5 conditions (1A-1E see below) x 2 screen position (left, right); we eliminated four double items in condition 1E (identical picture) as screen position was not relevant here. Each picture appeared in left and right screen position. Each item was a pair consisting of one picture and alternate picture. For example, a participant would see a picture of a man and a woman that place a lipstick horizontally and an alternate picture in one of the following conditions (see Figure 1):

1A. control (different picture): a koala
1B. critical (same object position, different gender agent): woman and woman place a lipstick horizontally
1C. critical (different object position, same gender agent): man and woman place a lipstick vertically
1D. control (different object): man and woman place a deodorant horizontally
1E. control (identical picture): man and woman place a lipstick horizontally

Figure 1: Picture examples of four similarity judgment items, each consisting of one picture and an alternate picture in condition 1A, 1B, 1C and 1D.



Categorization by Forced Choice

A set of 24 experimental items was created: 4 placement scenes (lipstick, deodorant, tube of toothpaste, binoculars) x 2 conditions (1B, 2B see below) x 3 screen position (left, middle, right). Each item was a triad consisting of one picture and two alternate pictures, i.e., participants see a scene (e.g., a man and a woman place a lipstick horizontally) and then two alternate pictures (see Figure 2):

2A. same object position, different gender agent: woman and women place a lipstick horizontally

2B. different object position, same gender agent: man and woman place a lipstick vertically

Figure 2: Picture examples of a forced choice triad item, consisting of one picture and two alternate pictures in condition 2A and 2B .

Linguistic Descriptions

Subjects received a form with eight pictures, depicting the four critical objects (lipstick, deodorant, tube of toothpaste, binoculars) both in horizontal and vertical manner and being placed by a man and woman or two women.

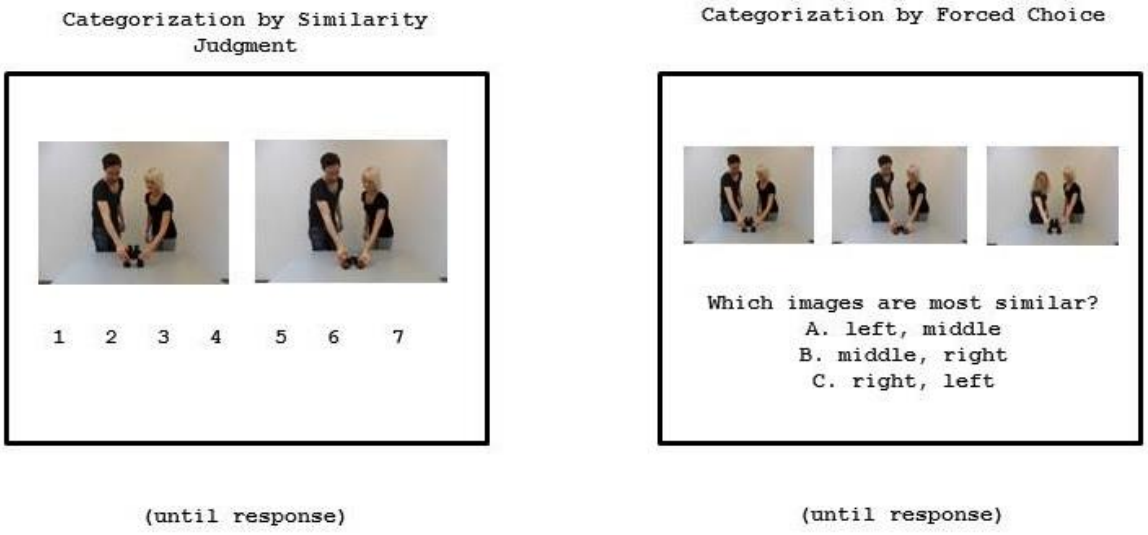
3.4 Procedure

Participants were tested in groups of 10-20 in a quiet computer room and they were instructed in their native language. All participants completed the following tasks in the following order: categorization of placement scenes by similarity judgment on computer; categorization of placement scenes by forced choice on computer; linguistic description task on paper. The order of administration of the different tasks was the same for all participants. The reason for this was that we first wanted to know about finer grained choices of participants with the similarity judgment task, before forcing them to choose among the different categorization options with the second task. The linguistic description task was given after the categorization tasks, to exclude possible effects of linguistic encoding before doing the categorization tasks. Implications of the order of administration are discussed in the Discussion.

Before embarking on the similarity judgment task, we gave participants two important instructions. First, we asked them to "Please note that the pictures you are about to see are *stills* taken out of a video. Press any key to watch such a video and still image". This was to ensure that participants perceived pictured events as "placement events" (lying, standing) and not as merely touching or holding an object. Second, we asked participants to vary their responses, in order to prevent them giving the same rating over and over again. These instructions were repeated before the categorization by forced choice task. For both the similarity judgment task and the forced choice task there were three practice trials.

Instructions for the similarity judgment part were: `You will see a number of picture combinations. They show people performing an action with an object or an animal that is looking into the camera. Please rate how similar you think the picture combinations are, by rating similarity on a scale from 1 till 7.´ Each picture set appeared on the screen with a scale from 1-7 presented below (see Figure 3). The scale ranged from "not similar at all" to "totally similar". Instructions for the categorization by forced choice part were: ´You will see three pictures at the same time. The pictures show people performing an action with an object. It is your task to select two out of three pictures that you think are most similar.´ Each picture triad appeared on the screen with the response options A (left, middle), B (middle, right) and C (right, left) presented (always in the same manner) below (see Figure 3). Each picture set or triad stayed on the screen, until the participant pressed a key to indicate his choice.

Figure 3: Example of a Similarity Judgment trial (left) and a Forced Choice trial (right)



After completing the categorization tasks participants received the linguistic description task. They were asked to describe the people, action and object depicted with a single sentence (eight in total), following an example. The example showed a picture of a woman closing curtains, described with the sentence "She closes the curtains". Participants were once again reminded that the pictures were *stills* taken out from videos to ensure they perceived the pictures as dynamic and not static events.

4. Results

4.1 Categorization by Similarity Judgment

Hypothesis 1 posed that German speakers would give scenes with similar object position higher similarity ratings than Spanish speakers, and conversely, Spanish speakers would give scenes where the gender of agents is the same higher similarity ratings than German speakers. For each participant, we calculated the mean similarity rating for each of the five conditions (1A-E, see Table 1). We treated this data as interval data. German and Spanish speakers gave almost identical similarity judgments for the five different conditions (see Table 1). A 2 language (German, Spanish) x 2 condition (1B,1C) mixed ANOVA with the mean ratings on condition 1B and 1C showed that there was no main effect of language, $F(1,43)=.442$, $p=.510$, $\eta_p^2=.10$. There was a main effect of condition, $F(1,43)=51.075$, $p<.001$, $\eta_p^2=.543$. On average, ratings for condition 1B ($M=4.34$) were lower than ratings for 1C ($M=5.33$). This means that picture sets where the gender of agents differed were perceived as less similar than picture sets where object position differed. There was no interaction between language and condition, $F(1,43)=.086$, $p=.771$, $\eta_p^2=.002$.

Table 1: Mean rating values and standard deviations in the Similarity Judgment task (scale 1-7)

		German Rating		Spanish Rating	
Condition	**Specification**	**Mean**	**St. Dev.**	**Mean**	**St. Dev.**
1A	**Control: Different Picture (Koala)**	1.13	0.32	1.07	0.23
1B	**Critical: Same Object Position, Different Gender Agent**	4.25	0.95	4.23	0.97
1C	**Critical: Different Object Position, Same Gender Agent**	4.91	0.95	5.05	0.67
1D	**Control: Different Object**	5.28	0.87	5.38	0.61
1E	**Control: Identical**	6.95	0.13	6.95	0.15

4.2 Categorization by Forced Choice

Hypothesis 2 posed that German speakers would make forced similarity choices on the basis of similarity in object position, whereas Spanish speakers would make their selection on the basis of similarity in the gender of agents. For each participant, we summed the number of trials where categorization choice was based on gender or positional aspects in pictures. A small percentage (3.1 %) of choices based on other aspects was left unconsidered. We then calculated the percentage of decisions based on gender and positional aspects in relation to the total number of choices made. We treated this data as interval data. German and Spanish speakers showed similar patterns across the different conditions (see Table 2). We ran an independent samples T-test on the average German and Spanish choices for condition 2A (a T-test on choices for condition 2B gave the same results). On average, the number of German choices based on similarity in object position (M=38.6%) was higher than Spanish choices based on similarity in object position (M=36.5%). This difference, 2.090, 95% CI [-19.746, 23.926] was not significant, $t(43)$=.193, p=.424, d=.058.

Table 2: Mean percentage of choices for gender and orientation aspects in the Forced Choice task

		German Choice		Spanish Choice	
Condition	**Specification**	**Mean %**	**St. Dev.**	**Mean %**	**St. Dev.**
2A	**Same Object Position, Different Gender Agent**	38.6	33.6	36.5	38.7
2B	**Different Object Position, Same Gender Agents**	61.4	33.6	63.5	38.7

4.3 Linguistic Descriptions

Finally we examined whether our expectations about cross linguistic differences in the descriptions of placement scenes by German and Spanish speakers are correct. In particular, we investigated whether German speakers described placement scenes where objects are in horizontal and vertical positions with the verbs *legen* [lay] and *stellen* [stand] and Spanish speakers with the verb *poner* [put] or *dejar* [leave in a place]. In addition, we examined whether Spanish speakers described placement scenes where there are female-only or mixed groups of agents with the pronouns *ellas* [they-feminine] and *ellos* [they-masculine] respectively and if German speakers employed the pronoun *sie* [they] in both cases.

For each participant, we counted the number of trials where the predicted form was used or whether another form was used. We calculated usage percentages of predicted forms in relation to the total number of trials. On average, German speakers used the expected verbs in 39.8 % (*SD*=42.4) and the Spanish speakers in 17.9 % (*SD*=25.0) of all trials. We also found that on average, German speakers used the expected personal pronouns in 19.3 % (*SD*=36.7) of trials and the Spanish speakers in 63.0 % (*SD*=45.6) of all trials. Since the linguistic descriptions did not match our predictions, we investigated the alternative descriptions speakers gave for horizontal/vertical and female-only/mixed group scenes.

First, we examined alternative descriptions for horizontal/vertical scenes. For German, alternative descriptions contained either a single verb (52.8%) or a verb with a specifier that indicated object orientation. The two most used single verbs were *halten* [hold] (25.8%) and *greifen* [grab] (21.3%). These verbs do not indicate object position. The two verbs most used in combination with a specifier were: *halten* (16.9%) and *greifen* (7.8%) with specifiers such as *waagerecht* [horizontal], *senkrecht* [vertical], *liegende(s)* [lying] and *stehende(s)* [standing]. For Spanish, alternative descriptions contained either a single verb (41.1%) or a verb with specifier that indicated object orientation. The two most used single verbs were *colocar* [place] (11.4%) and *coger* [take] (8.9%). These verbs do not indicate object position. The two verbs most used in combination with a specifier were: *sujetar* [hold] (20.3%) and *colocar* (15.8%) with specifiers such as *horizontalmente* [horizontal], *verticalmente* [vertical], *tumbado* [lying] and *de pie* [standing]. See Table 3 for an overview.

Table 3: Mean proportion of alternative verbs (plus specifiers) used to describe placement events in German and Spanish

Language	**Verb (+ specifier)**	**Mean**	**St. dev.**
German	Halten	25.8	1.60
	Greifen	21.3	1.16
	halten + waagerecht, senkrecht, liegende(s), stehende(s)	16.9	0.73
	greifen + waagerecht, senkrecht, liegende(s), stehende(s)	7.8	0.12
Spanish	sujetar + horizontalmente, verticalmente, tumbado, de pie	20.3	1.97
	colocar + horizontalmente, verticalmente, tumbado, de pie	15.8	1.42
	colocar	11.4	0.86
	coger	8.9	0.54

Second, we examined alternative descriptions for mixed/female only scenes. For German, alternative descriptions of men/women almost exclusively involved the nouns *Mann* [man] and *Frau* [woman] (91.5%) that specified the gender of the agent. For Spanish, alternative descriptions predominantly involved the nouns *hombre* [man] and *mujer* [woman] (61.7%), thus the gender of agents was specified.

The alternative linguistic strategies described above indicate that speakers were describing object position and gender of agents by other means than predicted. Therefore, we counted the number of (predicted and alternative) descriptions indicating object position and gender of agents for German and Spanish. We then calculated percentages in relation to the total number of descriptions given. Results are shown in Table 4.

Table 4: Mean percentages and standard deviations of German and Spanish descriptions containing indications of position and gender.

	Descriptions indicating Object Position		**Descriptions indicating Gender of Agents**	
Language	**Mean %**	**St.dev.**	**Mean %**	**St. dev.**
German	48.9	42.8	86.9	26.3
Spanish	54.3	41.9	86.4	34.3

A 2 language (German, Spanish) x 2 description (Object position, Gender agent) mixed ANOVA with the percentages of descriptions that described object position and the gender of agents showed there was no main effect of language, $F(1,43)=.093$, $p=.762$, $\eta_p^2=.002$. There was a main effect of description, $F(1,43)=22.298$, $p<.001$, $\eta_p^2=.341$. On average, participants described the gender of agents

(*M*=86.7%) more often than object position (*M*=51.7%). There was no interaction between language and description, $F(1,43)=.163$, $p=.688$, $\eta_p^2 =.004$.

5. Discussion and conclusions

The main goal of the present paper was to examine how German and Spanish speakers categorize the position of objects and gender of human agents in an under investigated type of motion event - placement events. Unlike previous studies, we investigated two conceptual domains, space and gender, within one study design. We examined how German and Spanish speakers categorized spatial and gender aspects in placement events when language was not explicitly offered or used when performing two types of categorization tasks. Also, we collected linguistic descriptions to examine whether the participants described the placement events following our linguistic predictions. This gave us an indication of their linguistic preferences, had they used language as a mental strategy during the categorization tasks.

This study yielded no support for Hypothesis 1 and 2 that posed categorization differences for German and Spanish speakers. We found that both German and Spanish speakers categorized spatial and gender aspects in placement events similarly as shown by similar ratings on a scale and by forced choice. In particular, results indicated that participants rather categorized based on similarities in the gender of human agents than on the basis of similarities in object position. What accounts for the lack of predicted differences in these behavioral patterns? One explanation is that participants did not employ language during the task. Therefore, cognitive differences guided by cross linguistic differences did not occur. Another explanation is that participants used language as a mental strategy during categorization, but used linguistic descriptions other than the ones predicted. This may have led to similar categorization preferences across groups. This explanation is supported by our analyses of linguistic descriptions. These analyses indicated that participants used alternative descriptions from the ones we predicted. Importantly, through both predicted and alternative descriptions both German and Spanish speakers indicated object position and gender of agents equally often.

Considering methodological limitations of this study, it is plausible that performance on the similarity judgment task affected performance on the categorization by forced choice task. Also, given the repeated changes in object position and gender of agents in the categorization tasks it is perhaps not surprising that both German and Spanish participants employed alternative linguistic strategies that marked object position and gender in the linguistic description task. Thus the possibility exists that the results on the categorization by forced choice and linguistic description task are not reflective of performance as compared with performance of participants that completed only one of the different tasks. Also, despite experimental instructions, it may be that some participants perceived the placement stimuli as static instead of dynamic events as indicated by their use of static verbs, such as “hold”, to describe the scenes, which may have affected their categorization responses.

Further, the investigation of two domains within one study brings along the issue of comparability. This applies to both linguistic instances as (visual) experimental manipulations. Considering the relevant linguistic instances, both verbs describing placement and personal pronouns describing human agents seemed comparable in the sense that they are basic and culturally significant productive lexical fields. Considering visual manipulations, the chosen cross linguistic differences involved manipulations that concerned both objects and human agents. For objects, we chose to change their position in space (horizontal vs. vertical); for human agents, we chose to replace the male agent by a female agent. The clothing (both agents were wearing the same black shirt) and posture of both agents were kept as similar as possible, yet aspects as body size, facial features, hair length and hair color differed between agents. The single orientation change of the object versus the multiple changes for the agent plausibly challenged the comparability of the different manipulations. In future work, a different manipulation applied to the same entity (the object) may provide a better case for comparison. For example, one could manipulate the size of objects following the German-Spanish cross linguistic difference in the use of augmentative suffixes to describe object size (Gooch, 1976; Butt & Benjamin, 2005; Lohde, 2006).

All in all, the results of this study cannot provide a definite answer to whether language preferences affect categorization in tasks that do not involve explicit use of language. However, they give the following insights into the matter. In a free encoding context, and for an unexplored type of motion event, placement events, no (spatial and gender) language effects on categorization could be found. A possible explanation is that participants employed alternative linguistic strategies when perceiving stimuli than the strategies we predicted. These alternative strategies may have resulted in similar descriptions of object position and gender of agents across languages, despite cross linguistic differences. Related studies in the motion domain have not reported that speakers employed alternative linguistic strategies from the ones predicted. In case the target linguistic constructions were produced during categorizing effects have been found (Gennari et al., 2002). However, when they were produced before categorizing (Papafragou, Massey, Gleitman, 2002; Cardini, 2010) or not at all (Gennari et al., 2002) no effects have been found. Considering our results and those of the studies just discussed, support in favor of a strong interpretation of the Sapir-Whorf hypothesis seems weak. Or, in words of Papafragou, Massey and Gleitman (2002): “the necessarily sketchy nature of language use assures that it will be at best a crude index of thought.”

Divergent results by Papafragou and Selimis (2010) for motion verbs may be explained by the fact that their task implicitly encouraged language use. In their case, participants may have actually employed the predicted linguistic forms. Divergent results by Athanasopoulos and Bylund (2013, 2014) may be explained as follows. Instead of examining manner of motion or different object positions resulting from motion their studies related more to the domain of time (aspect) than to motion. Lera Boroditsky and colleagues (2002; 2003) found effects for categorization that were related to gender encoded in object nouns and not to gender descriptions of human agents as in this study. It may be that the linguistic instances under investigation in these studies have a more pervasive effect on cognition than

those in our study. Such effects could be explained by different saliency of linguistic forms as determined by factors as their frequency of occurrence, word class and their (typical) position in sentences (Talmy, 2000; Ellis, 2004, 2008). Let us take frequency of occurrence as example. Berman and Slobin (1994) argue that the less frequent forms appear in the input the less salient they are. Spanish is a "pronoun-dropping" language, which means pronouns may be omitted from speech or writing unless they are needed for purposes of emphasis or contrast (Butt & Benjamin, 1998). Thus it is plausible that the Spanish object nouns used in the studies of Borodistky and colleagues appear more frequently in language input than the personal pronouns *ellos* [they-masculine] and *ellas* [they-feminine]. The difference in salience of critical forms in the separate studies would explain the discrepancy in results.

We argue that categorization tasks without explicit offering or production of language are not as informative as desired since it is not possible to determine what mental processes take place when participants perform such tasks (in line with Bylund & Athanasopoulos, 2014). Instead, tasks may be more informative when employed in the line of weaker interpretations of the Sapir-Whorf hypothesis (e.g. Thinking for Speaking, Slobin, (1996)). This would involve the employment of experimental contexts where subjects are presented with or urged to produce critical language right before or during categorization tasks. This may lead speakers to produce categorization differences or differences in other kinds of cognitive activities (e.g. memorizing).

In conclusion, this study indicates that cross linguistic differences do not affect how German and Spanish speakers categorize spatial and gender aspects in placement events. In a free encoding context speakers may (mentally) describe spatial and gender aspects through various linguistic means available in their language, resulting in similar descriptions across languages. This may rule out any cognitive preferences related to existing cross linguistic differences. All in all, our study does not support a strong interpretation of the Whorfian hypothesis and calls for a cautious approach when theorizing about language-categorization effects in contexts where no overt language is present.

Acknowledgements:
This work has been supported by the EU 7th Framework Programme Marie Curie Initial Training Networks grant Nr. 316748 under the project *Language and Perception*. I thank my collaborators, hosts and local assistants who enabled data collection at the test locations.

Endnotes

1. In the pilot study 12 participants were asked to consider 112 object combinations. They were asked how they would place the first object (e.g. a lipstick) in each combination onto the second object (e.g. a table). They responded by ticking a box on a five point scale that ranged from 1, "lying" to 3, "either way" to 5, "standing". For purposes of the current study we selected four object combinations that received an average rating of 3, "either way".

References

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., Riches, N., Thierry, G. (2015). Two Languages, Two Minds: Flexible Cognitive Processing Driven by Language of Operation. *Psychological Science,* 26, 518-526.

Athanasopoulos, P., Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science* 37(2), 286-309.

Berman, R., Slobin, D. (Eds.) (1994). *Relating events in narrative: A crosslinguistic developmental study (Vol. 1).* Hillsdale: NJ: Lawrence Erlbaum Associates.

Bosse, S., Papafragou, A. (2010). Spatial position in Language and Visual Memory: A Cross-Linguistic Comparison. *Proceedings from the 32nd Annual Meeting of the Cognitive Science Society*, Hillsdale, NJ: Erlbaum.

Boroditsky, L., Ham, W., Ramscar, M. (2002). What is universal in event perception? Comparing English & Indonesian speakers. *Proceedings of the 24th annual meeting of the Cognitive Science Society*, 136-141

Boroditsky, L., Schmidt, L., Philipps, W. (2003). Sex, syntax and semantics. In Gentner, D., & Goldin-Meadow, S. (Eds.), *Language in mind: Advances in the study of language and thought* (61-79), Cambridge: MA: MIT Press.

Butt, J., Benjamin, C. (1998). *A new reference grammar of modern Spanish*. New York: Oxford University Press.

Bylund, E., Athanasopoulos, P. (2015). Motion event categorization in a nativised variety of South African English. *International Journal of Bilingual Education and Bilingualism*, 18(5), 588-601.

Bylund, E., Athanasopoulos, P. (2014). Linguistic relativity in SLA: Toward a new research program. *Language Learning*, 64(4), 952-985.

Cardini, F. (2010). Evidence against Whorfian effects in motion conceptualization. *Journal of Pragmatics*. 42, 1442–1459.

Cartagena, N., Gauger, H. (1989). *Spanisch-Deutsch 1. Duden vergleichende Grammatiken*. Mannheim: Druckerei Krembel, Spreyer.

Ellert, M. (2010). Ambiguous pronoun resolution in L1 and L2 German and Dutch. Ph.D. dissertation submitted at Radboud University Nijmegen.

Ellis, N. (2004). The process of second language acquisition. In: VanPatten, B., Williams, J., Rott, S. (Eds.), *Form-meaning connections in second language acquisition* (pp. 49-76), Lawrence Erlbaum Associates Publishers, London, Mahwah: New Jersey.

Ellis, N. (2008). Usage-based and form-focused language acquisition: The associative learning of constructions, learned attention, and the limited L2 endstate. In P. Robinson & N. Ellis (Eds.): *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 372-405). New York, Abingdon: Routledge.

Evans, N. 2011. Semantic typology. In J.J. Song (Ed.), *The Oxford handbook of linguistic typology* (504-533). Oxford: Oxford University Press.

Fausey, C., Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin and Review*, 18(1), 150-157.

Gardner, H. (1985). *The minds new science*. A history of the cognitive revolution. New York: Basic Books, Inc. Publishers.

Gennari, S.P., Sloman, S.A., Malt, B.C, Fitch, W.T. (2002). Motion events in language and cognition. *Cognition* 83, 49-79.

Gentner, D., Goldin-Meadow, S. (2003). Whither Whorf. In Gentner, D. & Goldin-Meadow (Eds.): *Language in Mind. Advances in the Study of Language and Thought.* Cambridge, London, Massachusetss: MIT Press.

Gooch, A. (1967). *Diminutive, augmentative and pejorative suffixes in modern Spanish. A guide to their use and meaning*. Oxford: Pergamom Press.

Hae in Park, L., Ziegler, N. (2014). Cognitive shift in the bilingual mind: Spatial concepts in Korean-English bilinguals. *Bilingualism: Language and Cognition,* 17(2), 410-430.

Hellinger, M., Bußmann, H. (Eds.) (2001, 2002, 2003). *Gender across languages: The linguistic representation of women and men.* Amsterdam: John Benjamins Publishing Company.

Kopecka, A., Narasimhan, B. (Eds.) (2012). *Events of putting and taking: A crosslinguistic perspective*. Amsterdam: Benjamins.

Lamers, M., Jansma, B., Hammer, A., Münte, T., (2008). Differences in the processing of anaphoric reference between closely related languages: neurophysiological evidence. BMC Neuroscience, 9:55.

Lohde, M. (2006). *Wortbildung des modernen Deutschen: ein Lehr- und Übungsbuch.* Tübingen: Gunter Narr.

Malt, B.C., Majid, A. (2013). How thought is mapped into words. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4, 538-597.

Nosofsky, R. (1986). Attention, Similarity and the Identification-Categorization Relationship. *Journal of Experimental Psychology: General*, 115(1), 39-57.

Papafragou, A., Massey C., Gleitman, L. 2002. Shake, rattle, 'n' roll: the representation of motion in language and cognition. *Cognition* 84, 189-219.

Papafragou, A., Selimis, S. (2010). Event categorisation and language: A cross-linguistic study of motion, *Language and Cognitive Processes*, 25(2), 224-260,

Philipps, W., Boroditsky, L. (2003). Can quirks of grammar affect the way you think? Grammatical gender and object concepts. *Proceedings of the 25th annual meeting of the cognitive science society*, 928-933.

Slobin, D.I. 1996. From "thought and language" to "thinking for speaking." In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity. Studies in the social and cultural foundations of language*, vol. 17 (70-96). Cambridge: Cambridge University Press.

Talmy, L. (1975). Semantics and syntax of motion. In: John Kimball (ed.), *Syntax and semantics, 4*, (pp. 181-238), New York: Academic Press.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description, vol. 3: Grammatical categories and the lexicon* (pp. 36-149), Cambridge: Cambridge University Press.

Talmy, L. (2000). Toward a cognitive semantics: Typology and process in concept structuring (Vol. II), MA: MIT Press.

Trueswell, Papafragou (2010). Perceiving and remembering events cross-linguistically: evidence from dual-task paradigms. *Journal of Memory and Language*, 64, 64-82.

Van Bergen, G., Flecken, M. (2017). Putting things in new places: Linguistic experience modulates the predictive power of placement verb semantics. Journal of Memory and Language, 92, 26-42.

Whorf, B., Bissel Carroll, J. (1956). *Language, Thought and Reality. Selected Writing of Benjamin Lee Whorf*. MIT Press.

Wolff, B., Holmes, K. (2010). Linguistic relativity. *WIREs Cognitive Science,* 2(3), 253-265.

Paper 2:

Dietha Koster, Kenny Coventry, Teresa Cadierno

Standing, lying, male or female? The effect of language on recognition memory in placement events for L1 and L2 speakers of German and Spanish.

Standing, lying, male or female? The effect of Language on Recognition Memory of Placement Events for L1 and L2 speakers of German and Spanish

Dietha Koster, Kenny Coventry, Teresa Cadierno

Abstract: Language and perception are essential cognitive systems, yet how do changes in one system affect the other? In this paper we consider adult second language (L2) learning as a study case. We take linguistic variation in German and Spanish in the description of placement events (e.g. *They put the binoculars on the shelf*) as a starting point. In placement events, German marks object position by its verbs *legen* [lay] and *stellen* [stand] whereas the Spanish *poner* [put] does not. However, Spanish marks gender of agents by *ellos* [they-masculine] and *ellas* [they-feminine] whereas the German *sie* [they] does not. In a computer-based memory experiment, we presented native (L1) and L2 speakers of German and Spanish with language and pictures describing and showing placement events. In critical memory trials, we presented speakers with pictures where we changed the position of objects or facial gender of agents. First, we ask whether cross linguistic differences affect how L1 German and Spanish speakers perceive object position and facial gender of agents in placement events. Second, we investigate whether the L2 affects perception when a German learns Spanish and a Spaniard learns German. Third, we examine whether L2 proficiency mediates potential perceptual effects. We compare recognition memory accuracy and reaction time (RT) data of L1 speakers and lower and higher proficient L2 speakers of German and Spanish. Results show three things. First, not gender, but spatial language affects recognition memory for object position in L1 speakers and L2 speakers. Second, L2 German speakers´ recognition memory accuracy for object position is as good as L1 German speakers´ memory. Third, this memory effect is not mediated by L2 proficiency. We discuss implications of these results for theory on recognition memory in relation to language within L1 and L2 speakers.

Keywords: Thinking for Speaking, language processing, recognition memory, L2 proficiency

Highlights:
- L1 German speakers have better recognition memory for object position than L1 Spanish speakers
- When German learners of L2 Spanish perform the memory task in Spanish, their recognition memory for object position is similar to L1 Spanish speakers´ memory
- When Spanish leaners of L2 German perform the memory task in German, their recognition memory for object position is as good as L1 German speakers´ memory

1. Introduction

To understand the world around us, our brain must analyze incoming patterns of information and compare them with information already stored in memory. For example, imagine seeing a picture of a man placing binoculars on a shelf. Later, you encounter a second, similar picture and are asked to compare it with the first one. Will the language one speaks affect what is remembered about the first picture and the latter comparison as a result? There is evidence that language can indeed affect what we attend to and what is remembered about images. For example, Gennari et al. (2002) found that English motion verbs lead English speakers to have better memory for manner of motion (e.g. clamber, strid, creep) in motion events as compared with Spanish speakers. English motion verbs normally encode manner of motion whereas in Spanish, manner of motion is expressed optionally through adverbial phrases (e.g. *entra caminando* [(he) enters walking]. Boroditsky, Schmidt and Philipps (2003) showed that Spanish and German speakers´ memory for object-name pairs (e.g. *apple-Patricia*) was better for pairs where the gender of the proper name was consistent with the grammatical gender of the object name in their L1. The attribution of grammatical gender to the particular objects was reversed for Spanish and German. The effect was shown even though participants performed the memory task in English.

The studies above relate to a research tradition investigating the Sapir-Whorf or linguistic relativity hypothesis (Whorf, 1956) or Slobin´s (1996) Thinking for Speaking (TFS) hypothesis. The main idea of both hypotheses is that speakers of different languages describe the world differently and therefore must perceive the world in a different way. While the Whorfian hypothesis poses that language effects on nonlinguistic cognition (e.g. perception, memory) exist even when language is not overtly used, the TFS hypothesis poses effects appear when individuals are engaged in language-driven activities. Research into linguistic relativity and TFS questions has experienced a revival in the last decades (Gumperz & Levinson, 1996; Gentner & Goldin-Meadow, 2003). Linguistic relativity can be said to comprise a family of related proposals on the relation between languages and thought (for discussion see Wolff & Holmes, 2010). The literature is unresolved about if and in which contexts the Sapir-Whorf hypothesis holds (Trueswell & Papafragou, 2010; Fausey & Boroditsky, 2011) and whether effects are driven by language at encoding or retrieval (Feist & Gentner, 2007). In recent years, Whorfian and TFS research has started to focus increasingly on bilingual speakers (Cook & Basetti, 2011; Han & Cadierno, 2011; Pavlenko, 2014). In this paper we aim to address two Whorfian or TFS research questions in relation to bilingualism.

A first important theoretical question is whether bilinguals´ L2 affects their memory of motion events. There is little work available on this topic, as pointed out by Filipovic (forthcoming). A study by Filipovic (2011) suggests that L2 speakers´ recognition memory may be affected by how they express manner of motion. She examined how Spanish-English bilinguals described and remembered motion events as compared with L1 controls. The L2 results showed that English and Spanish descriptions by bilingual participants adhered to a lexicalization pattern that is acceptable in both languages, which was

Spanish in this case. Consequently, Filipovic found that bilinguals´ memory results were in line with those of L1 Spanish speakers. There are also few studies that employ a bidirectional design. Filipovic´s (forthcoming) study on variation in the expression of intentionality in causation events is an exception, as she examined both proficient English learners of L2 Spanish as well as Spanish learners of L2 English (and L1 controls). Spanish speakers distinguish between intentional and non-intentional events whereas English lexicalization patterns often leave intentionality underspecified. This study shows that L2 learners´ recall memory relied on their L1 even when they used their respective L2s exclusively in the experiment. Filipovic´s (2011; forthcoming) studies show that language may have different effects on recognition and recall performance. With respect to recognition memory, we may predict that L2 learners´ recognition memory will be affected by the language they employ.

Another important question is whether L2 proficiency mediates potential L2 effects on memory of motion events. Francis (2005) has pointed out there is little work examining cognitive processes (e.g. memorizing) in bilinguals across different proficiency levels. Central in this paper are so called "subordinate" bilinguals (Weinreich, 1953) with lower and higher L2 proficiency levels. A typical subordinate bilingual is a classroom learner, who learns the L2 while an L1 system is already in place. Research has shown that L2 learners with higher levels of L2 proficiency process sentences almost as automatic as L1 speakers whereas speakers at lower levels of L2 proficiency do not (Rossi, Gugler, Friederici & Hahne, 2006; Van Hell & Tokowicz, 2010; Foucart & Frenck-Mestre, 2013). This suggests that L2 processing imposes a higher cognitive load on L2 learners with lower L2 proficiency as compared with L2 learners with higher L2 proficiency and L1 speakers. It has been shown that a higher cognitive load caused by performing multiple tasks simultaneously impairs both recall and recognition (e.g. Hicks & Marsh, 2000; Naveh-Benjamin, Craik, Guez & Dori, 1998). Considering these findings on L1 and L2 processing, cognitive load and memory impairment, we may predict that L2 speakers´ recognition memory will be worse than that of L1 speakers; and L2 learners with lower L2 proficiency will have a worse recognition memory than L2 learners with higher L2 proficiency.

The present bidirectional study provides novel empirical evidence that may help unveil language-specific effects on memory. In this paper, we take variation in German and Spanish in the description of placement events (e.g. *He puts the binoculars on the shelf*) as a starting point. German and Spanish vary in their typical ways of expressing the position of an object in space. German verbs mark horizontal (*legen*, [lay]) or vertical (*stellen*, [stand]) object position whereas the typical Spanish placement verb (*poner* [put]) does not. German and Spanish also differ as to how they express gender of multiple agents. Spanish personal pronouns mark gender for female groups (*ellas*, [they-feminine]) and male or mixed groups (*ellos*, [they-masculine]) whereas the German pronoun *sie* [they] does not. We investigate native (L1) speakers of German and Spanish, Spanish learners of L2 German; and German learners of L2 Spanish. We investigate both lower and higher proficient L2 learners. We present participants with language and pictures that describe and show people placing objects. The fact that we present subjects with language positions this study in the TFS line of research. To our

knowledge, there are no studies into L2 memory that consider variation in the expression of (motion in) space and gender in placement events, let alone within one single investigation.

Our aim is to examine whether language affects recognition memory in L1 and L2 speakers; and whether L2 proficiency mediates potential effects for L2 speakers. We compare accuracy and RT data of German and Spanish speakers to critical recognition memory trials. In these trials we altered object position and facial gender of agents with respect to preceding prime pictures. We compare responses between L1 speakers; we compare L1 with L2 speakers and we compare three different proficiency groups for each language. In the following we discuss the theoretical background of this study in more detail. We first provide information about our test bed: variation in the description of space and gender in placement events. Thereafter, we review empirical findings for language effects on memory for L1 and L2 speakers. Finally, we discuss the role of L2 proficiency in relation to language processing and potential effects for recognition memory. Our review will lead to the formulation of testable hypotheses.

2. Theoretical Background

2.1 Test Bed: Space and Gender in Descriptions of Placement Events

The starting point of TFS research is that speakers of different languages encode the world differently. Thus a critical tool is cross-linguistic comparison. Motion events have received considerable interest cross linguistically following Talmy´s typology of motion events (Talmy, 1975, 1985, 2000). In this paper, we focus on a particular type of motion event, which is a "placement event". In placement events, an agent or multiple agents cause(s) an object to move to a certain location (e.g. *They put the binoculars on the shelf).* Research suggests that the basic components in placement events are: Figure (what is moved), Agent (the causer of the movement), Ground (the location where an object is placed), Causation (what triggers the placement), Motion (the act of moving), and Path (the trajectory followed by the Figure) (Talmy, 1985; Jackendoff, 1990).

Cross linguistic research has revealed differences in how languages describe the position of Figure objects in placement events (Kopecka & Narasimhan, 2012). For example, German and Spanish verbs differ in their expression of spatial position of Figure objects in placement scenes. In German, typically, one would not use a general verb like ´put´, but a verb that either marks horizontal (*legen*, [lay] or vertical (*stellen*, [stand]) orientation of the Figure object being placed (Fagan, 1991; Berthele, 2012). Spanish employs verbs (*poner*, [put]; *dejar* [leave in a place] that do not mark object orientation (Ibarretxe-Antuñano, 2012, Cadierno et al., 2016). See Figure 1 for examples of two placement events and German and Spanish descriptions of these events. Empirical work shows that variation in the semantics expressed by placement verbs may affect how fast spakers of different languages move their eyes towards Figure objects when reading placement verbs (Van Bergen & Flecken, 2017).

Figure 1: Picture and German and Spanish sentence examples of two placement events with horizontal (left) and vertical (right) Figure object position and Agents with male and female gender (left) and female gender (right)

Language	**Sentence Description**	**Sentence Description**
German	**Sie legen** das Fernglas auf den Tisch. *They lay the binoculars on the table.*	**Sie stellen** das Fernglas auf den Tisch. *They stand the binoculars on the table.*
Spanish	**Ellos ponen** un par de binoculares en la mesa. *They(masculine) put the binoculars on the table.*	**Ellas ponen** un par de binoculares en la mesa. *They(feminine) put the binoculars on the table.*

German and Spanish also differ as to how they describe the gender of multiple Agents in placement events through personal pronouns. In case there are multiple agents performing a placement action, in Spanish we must mark whether the agents are female, *ellas* [they-feminine], or whether it is a mixed or male-only group, *ellos* [they-masculine] (Butt & Benjamin, 1998). We note that in Spanish, these pronouns are expressed for purposes of emphasis of contrast. German employs a single pronoun *sie* [they] to describe both female and male constellations (Drosdowski, 1995). See Figure 1 for examples of two placement events and German and Spanish descriptions of these events. We know of little empirical work on how speakers perceive gender expressed through personal pronouns. However, Lamers et al. (2008) showed that event-related brain potentials (ERPs) of Dutch and German speakers differed when processing male, female and neutral pronouns. They related these differences to language specific characteristics concerning the Dutch neutral pronoun *het* [it].

Following the TFS hypothesis we may predict the following. If speakers encounter the linguistic forms described above when perceiving placement events with multiple actors, German will direct its L1 and L2 speakers to attend to object orientation. Attention is the process of focusing psychological resources

to enhance perception, performance and mental experience (Bernstein, 2011). In contrast, Spanish will point its L1 and L2 speakers to attend to the gender of agents.

2.2 Recognition Memory in L1 and L2

Do distinctions languages make affect recognition memory for visual images in L1 and L2 speakers? Recognition memory is the ability to recognize previously encountered events, objects or people. Behaviorally, recognition memory can be measured through accuracy of retrieval in memory tasks. A memory task typically involves the presentation of pictures and/or words or sentences that have to be retrieved later on. Heredia and McLaughlin (1992) have argued that task requirements seem to be decisive in shaping the form that bilingual memory may take. Our study addresses this issue by showing new empirical evidence on a fine-grained task.

A number of empirical studies have provided insight into recognition memory for pictures in relation to L1 and L2. We focus on TFS studies on spatial (including motion) and gender language. Considering L1 and the domain of (motion in) space, Feist and Gentner (2007) found that spatial prepositions lead English speakers´ recognition memory to be accurate after reading spatial prepositions. In particular, their participants (falsely) recognized stronger versions of spatial relationships (e.g. a block on top of a building) as identical to preceding weaker ones (e.g. a block on the edge of a building) after reading spatial prepositions (e.g. "*The block is on the building*."). However, Coventry et al. (2010) found no such results in a study that employed a similar experimental task as Feist and Gentner (2007). They investigated whether English-Spanish cross linguistic differences in the expression of containment and support affected recognition memory, but found no effects. Bosse and Papafragou (2010) also found no effects of language on memory in a study on the German verbs *sitzen* [sit] and *liegen* [lie]. On the contrary, Billman et al. (2000) and Gennari et al. (2002) showed that speakers had better memory for manner of motion in case they heard or verbalized L1 verbs that encode manner of motion as compared with speakers whose language normally does not encode manner of motion. In the domain of gender, Boroditsky, Schmidt and Philipps (2003) showed that Spanish and German speakers´ memory for object-name pairs (e.g. *apple-Patricia*) was better for pairs where the gender of the proper name was consistent with the grammatical gender of the object name in their L1. Overall, it seems that number of L1 studies with evidence in favor of better memory performance through language is as high as the number of studies with evidence for worse (Feist and Gentner, 2007) or null (Bosse & Papafragou, 2010; Coventry et al., 2010) memory performance. We know of only one study on recognition memory in L2 speakers in the domain of motion. Filipović (2011) found that English-Spanish bilinguals´ descriptions of motion aligned with Spanish lexicalization patterns and L1 Spanish speakers´ recognition memory. This study suggests that L2 learners´ recognition memory may be affected by the language they employ.

We know of no studies into whether L2 proficiency mediates potential L2 effects on memory for pictures. However, L2 research indicates processing differences for L1 speakers and higher and lower proficient L2 learners at both word level (Kroll & Stewart, 1994; Dufour & Kroll, 1995) and sentence

level (for reviews see Van Hell & Tokowicz, 2010; Foucart & Frenck-Mestre, 2013). For example, Rossi et al. (2006) measured L1 Italian and L1 German speakers´ and lower and higher proficient L2 Italian and German learners´ event-related potentials (ERPs) during sentence processing. Sentences contained word category violations and morphosyntactic agreement violations. High-proficiency learners in both languages showed the same ERP components as L1 speakers for syntactic violations. In addition, the timing of processing steps was equivalent to that of L1 speakers, though some amplitude differences were present. Low-proficiency L2 learners, however, showed qualitative differences in the agreement violation characterized by an anterior negativity (LAN). A LAN is characterized by a negative-going wave that peaks around 200 milliseconds or less after the onset of a stimulus. They also showed quantitative differences reflected in a delayed P600 in every violation condition, indicating more uncertainty and problems during syntactic reanalysis. The P600 is a peak in electrical brain activity that occurs when processing grammatical errors. All in all, these results suggest that L2 processing may differ qualitatively and quantitatively at lower and higher levels of L2 proficiency. Importantly, L2 processing in learners at lower levels of L2 proficiency may impose a higher cognitive load as compared with L2 learners with higher L2 proficiency and L1 speakers (Abu-Rabia, 2003; Ransdall, Arecco & Levy, 2000). It has been shown that a higher cognitive load or divided attention impairs both recall and recognition (e.g. Hicks & Marsh, 2000; Naveh-Benjamin, Craik, Guez & Dori, 1998). For example, Hicks and Marsh (2000) made participants perform tasks as randomly generating numbers (1-10) or letters (A, B, C, D) or adding up sequences of numbers (1-9) during a recognition memory task. They found that memory performance, both in terms of accuracy and RT, was impaired in all these conditions in comparison with a control group that performed no additional tasks. Considering findings on L2 processing, cognitive load and memory impairment, we may predict that L2 speakers´ recognition memory will be worse than that of L1 speakers; and L2 learners with lower L2 proficiency will have worse recognition memory than L2 learners with higher L2 proficiency.

The reviewed studies above enable us to extend predictions made so far. We predicted that the German *legen/stellen* [lay/stand] distinction would make German speakers attend to object orientation and the Spanish *ellos/ellas* [they-masculine/-feminine] distinction would make Spanish speakers attend to the gender of agents. In extension, based on findings of Billman et al. (2001), Gennari et al. (2002) and Filipovic (2011) we may predict that attention guided by language will lead to better recognition memory of spatial and gender aspects in pictures depicting placement events. Further, following research on L1 and L2 processing, cognitive load and memory impairment we predict that L2 speakers´ recognition memory will be worse than that of L1 speakers and that L2 learners with lower L2 proficiency will have worse recognition memory than L2 learners with higher L2 proficiency.

2.3 Hypotheses

We set up the following scenario to test the predictions described above. L1 and L2 speakers of German and Spanish read sentences describing placement scenes (as in Figure 1). In German, these

sentences contain the verbs *legen/stellen* [lay/stand]; in Spanish, the sentences contain *ellos/ellas* [they-masculine/-feminine]. Simultaneously they see pictures of placement scenes. In a subsequent memory task, we change object orientation and facial gender of agents in these pictures and probe their recognition memory. The hypotheses we aim to address are as following:

1. L1 German speakers will have a better recognition memory for object orientation than L1 Spanish speakers and conversely, L1 Spanish speakers will have a better recognition memory for facial gender of agents.

2. When Spanish learners of L2 German and German learners of L2 Spanish perform the task in their L2, their recognition memory will be affected by the spatial or gender property encoded in the L2.

3. L2 learners of Spanish and German will have a recognition memory that is worse than that of L1 speakers of these languages respectively.

4. L2 proficiency will mediate the effect of the L2 on recognition memory. Lower proficient L2 learners will have worse recognition memory for the property encoded in their L2 than higher proficient L2 learners.

3. Method

L1 and L2 Participants

We recruited 27 native speakers of German at the University of Bremen (Germany) and 27 native speakers of Spanish at the University of Seville (Spain). Second, we recruited German learners of L2 Spanish in Germany and Spanish learners of L2 German in Spain. We found the 123 Spanish learners of L2 German through the German department at Seville University and Granada University and the *Goethe Institut* Granada (Spain). The 141 German learners of L2 Spanish were recruited through *Instituto Cervantes* in Bremen and Berlin; *Sprachenzentrum* Humboldt Universität Berlin; *VHS* Berlin Mitte; and *Sprachenzentrum* Universität Münster (Germany). Participant details are provided in Table 1. Participants received a nominal fee for their participation.

Table 1: L1 and L2 participant details

Language	**N**	**% Female**	**Mean Age**
L1 German	27	51.9	23.5
L2 German A2	58	60.3	21.9
L2 German B1	45	77.3	20.5
L2 German B2+	20	55.0	21.5

Language	N	% Female	Mean Age
L1 Spanish	27	74.0	21.3
L2 Spanish A2	31	71.0	23.1
L2 Spanish B1	52	71.2	22.0
L2 Spanish B2+	58	72.4	24.5

L2 Proficiency

L2 participants reported their level of proficiency in terms of the Common European Framework of Reference (CEFR). CEFR is a framework that has been established by the Council of Europe as a way to standardize the language learning achievements of language learners across Europe (Council of Europe, 2001). The CEFR has six levels of assessment: A1 and A2 (basic user), B1 and B2 (intermediate user) and C1 and C2 (proficient user).The reason to choose CEFR level as proficiency measure was that it was comparable across the Spanish and German context and that all institutes where we recruited students placed their students at different L2 proficiency levels following the CEFR. We asked students to report the level at which they were currently placed or if they did not study currently, the last level attained (also see Thompson, 2015 on the validity of (self-reported) CEFR L2 proficiency levels).

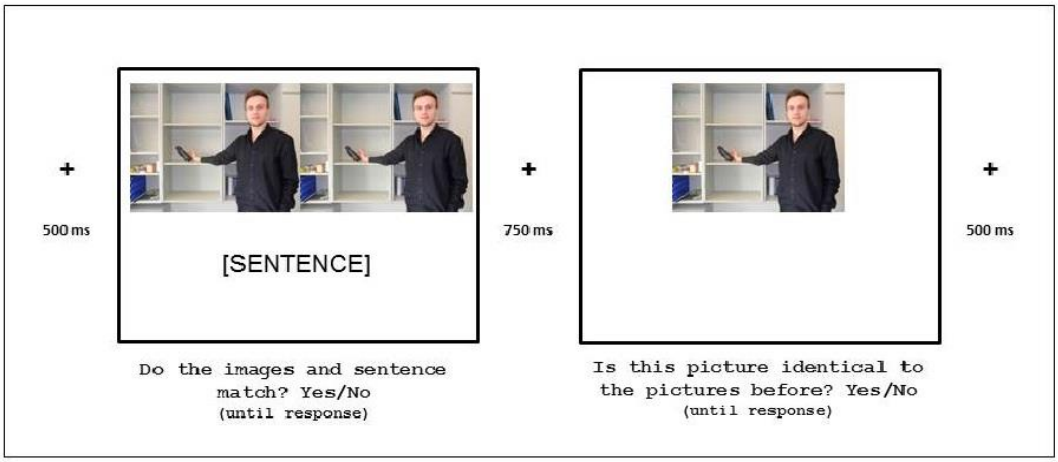
Despite numerous data collection efforts (several visits to Germany and Spain with multiple participation options in terms of test location and time) we ended up with unequal numbers of participants for different CEFR levels. For analyses we divided L2 learners over three groups: beginners, who reported A2 level; intermediate learners, who reported A2-B1, B1 or B1-B2 level; and advanced learners, who reported to have B2 level or higher. There was a considerable difference in the size and composition of the B2+ groups for L2 German and L2 Spanish. The B2+ German group (N= 20) contained learners at the following self-reported CEFR levels: 12 B2; 3 B2-C1; 3 C1; 2 C2. The B2+ Spanish group (N=58) contained learners at the following self-reported CEFR levels: 27 B2; 9 B2-C1; 18 C1 and 5 C1-C2 learners.

Tasks and measures

We employed a task adapted from Coventry et al. (2010). Subjects completed a computer-based task with two phases per trial, as depicted in Figure 2. In the first phase, participants were asked to verify whether a sentence matched two identical pictures. The purpose of the task was to ensure that participants read the sentence and studied the pictures. In the second trial phase, participants were asked whether a presented picture was identical to the preceding one(s). The purpose of this task was to determine whether participants recognized changes in pictures or not and critically whether recognition success is affected by language at encoding and L2 proficiency. We obtained subjects´ answers (Yes or No) and their RTs. Responses were given by pressing the Q and P key with the right and left index fingers. We created a right- and left-handed version of the experiment with reversed Yes/No response

options. The reason for having two identical pictures in the first phase of the task instead of one picture was as follows. "They"-sentences had to refer to two agents at least in order to use the Spanish *ellos/ellas* [they-masculine/feminine] to describe the scene. However, considering the second trial phase, we needed only one placement scene as referent. Therefore, we presented subjects with two identical pictures showing two identical agents. Admittedly, the result was somewhat artificial, yet pilot participants indicated no difficulties in judging the linguistic forms in connection to the two pictures.

Figure 2: Example of a two-phased experimental trial



Materials

We used language (sentences) stimuli and picture stimuli. The critical language stimuli in trial phase 1 were 48 sentences describing placement scenes (e.g. *Sie stellen das Fernglas auf das Regal* 'They put the binoculars on the shelf'). The sentences described four critical placement scenes, which were: man puts binoculars on shelf; man puts glue stick on plate; woman puts flashlight on piece of paper; woman puts lipstick on cutting board. Through a pilot study, we chose object combinations with no "typical" placing orientation of the figure object (e.g. a plate is typically placed horizontally and not vertically onto a table, thus this object combination was not appropriate for purposes of this study)[1]. This was to ensure that in German, participants would only be primed by the verbs *legen*/*stellen* [lay/stand] and not by a preferred placing orientation. We created two prime sentences and one "neutral" sentence (in German and Spanish) describing the four scenes. The neutral sentence did not indicate object position. It was created to examine for German speakers whether any language effects on memory were due to the German placement verbs only or whether they extended to neutral sentences. German and Spanish sentence examples with English translations are given in Table 2. Each sentence appeared in four different sentence-picture conditions (3 sentence (prime1, prime2, neutral) x 4 picture (orientation change, gender change, object disappears, identical) x 4 placement scene (binoculars, glue stick, flashlight, lipstick). The different (sentence-) picture conditions are discussed below.

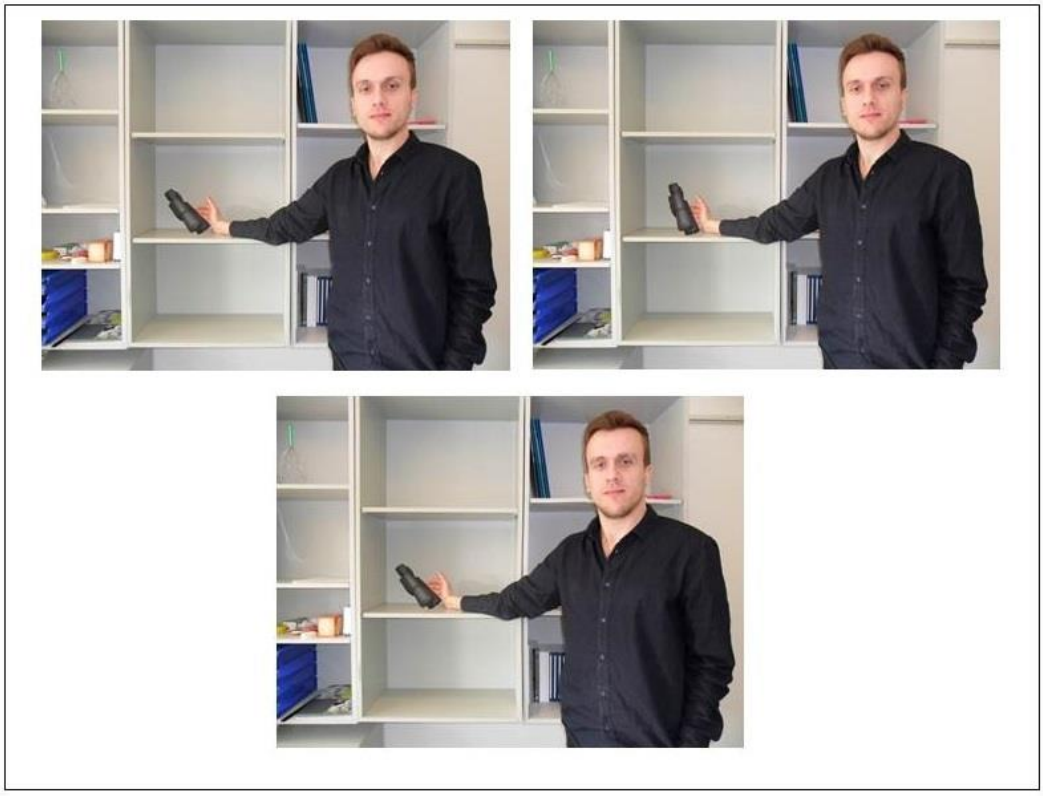
Table 2: Examples of two prime sentences and one "neutral" sentence for German and Spanish, with English translations.

Sentence Type	**German**	**Spanish**
Prime1	**Sie stellen das Fernglas auf das Regal.** They stand the binoculars on the shelf.	**Ellos ponen un par de binoculares en un estante.** They(-masculine) put the binoculars on the shelf.
Prime2	**Sie legen das Fernglas auf das Regal.** They lay the binoculars on the shelf.	**Ellas ponen una barra de labios en una tabla de cortar.*** They(-feminine) put the lipstick on the cutting board.
Neutral	**Es gibt Männer, Ferngläser und Regale.** There are men, binoculars and shelves.	**Hay hombres, pares de binoculares y estantes.** There are men, binoculars and shelves.

* Whereas German verbs were used within the same placement scenes; Spanish pronouns differed between placement scenes.

The critical picture stimuli in trial phase 1 and 2 were 16 color pictures. The standard prime picture (trial phase 1) showed an agent placing an object, with the object held at an angle of 45 degrees. Each prime picture was followed by a recognition picture (trial phase 2) that appeared in one of four conditions. The recognition picture was either identical or the object disappeared (control conditions); or we changed the object angle or facial masculinity (critical conditions). See Figure 3. The object angle was changed into 65 or 25 degrees; congruent with the German verb in trial phase 1. Facial masculinity was increased or decreased by 160% (by means of Psychomorph software, see De Bruine et al., 2006); incongruent or congruent (50/50) with the Spanish noun in trial phase 1. 48 Filler sentences and pictures described and showed agents performing other actions with other objects (e.g. *They open the door of the office*). In total the experiment had 96 trials.

Figure 3: Example of a standard picture (object 45 degrees) (left); a picture with orientation change (object 65 degrees) (right); and a picture with a gender change (increased facial masculinity) (below)



Design

Each participant completed a 3 (sentence type: prime1, prime2, neutral) x 4 (picture: orientation change, gender change, object disappears, identical) computer-based picture priming experiment. Filler items were designed for control purposes and were not analyzed.

General procedures

Participants were tested in groups of 10-20 in a computer room. L1 and L2 German speakers were given instructions in German; L1 and L2 Spanish speakers were instructed in Spanish. The experimental instructions, employing a cover story so participants would not guess the hypotheses, were as follows: ´In this experiment we investigate binocular vision. Binocular vision is sight where one uses both eyes. Sometimes one sees different things with the left and right eye, and we want to find out how this works. You will see a number of phrases and pictures that describe and show people performing an action with an object. First, you will see two identical pictures and a sentence. You are asked to verify whether the pictures match the sentence or not. Give your answer by pressing Q(Yes) or P(No). Second, you will see a single picture. You are asked to verify whether this picture is identical to the pictures you just saw. Please do not take the number of pictures into account (the fact that there

were two images first, and then one), but what is shown on the pictures. Differences can be small or large. Give your answer by pressing Q(Yes) or P(No). Please concentrate during the experiment and take the time you need to complete it. Keep your fingers at the Q and P keys during the whole experiment.´ There were three practice trials to familiarize participants with the task. The order of trials was randomized.

L2 procedures

The meaning of the Spanish *ellos/ellas* [they-masculine/-feminine] is typically taught in Spanish books at A1-level (e.g. see *Gente.hoy 1*, A1-A2). Therefore, we assumed that all German L2 Spanish learners knew the meaning of the pronouns, which meant these learners did not receive a linguistic instruction before following the general procedures. In contrast, we found it is not standard to teach the meaning of the German *legen*/*stellen* [lay/stand] in L2 German text books (but see De Knop & Perez (2014) for some textbook examples). To ensure that all L2 German learners knew the critical forms before they did the memory experiment, they received a linguistic instruction before following the general procedures (see Appendix 1). We expected that the instruction would provide new information for lower proficient L2 learners and consolidation of knowledge for higher proficient L2 learners. First, L2 German learners completed an exercise where they had to indicate for 18 German verbs whether they indicated horizontality, verticality or neither of these options (pretest). Next, they received a linguistic instruction where the meaning of the *legen/stellen* [lay/stand] verbs was explained (5 minutes). After the instruction they filled out the exercise again (posttest, order of items randomized) to check whether the training had facilitated learning. The average score (N= 123) on the pretest was *M*=60.0 % (*SD*=21.3%, min=11.1%, max=100%) and the average score on the posttest was *M*=90.8% (*SD*=17.2%, min=22.2%, max=100%).

4. Results

In this section we discuss accuracy and RT scores for the recognition memory task. L1 and L2 speakers indicated whether a presented picture was identical as compared with preceding pictures. We obtained their answers (Yes or No) and their RT. Below, we first compare accuracy data of L1 German and Spanish speakers; then we compare L1 and L2 German speakers, followed by L1 and L2 Spanish speakers. We also compare RT data to exclude the option of accuracy differences through longer study times. For L1 and L2 German speakers we compare accuracy results for trial with spatial language and trials without spatial language. Means and standard deviations for measures reported in the analyses can be found in Table 5.

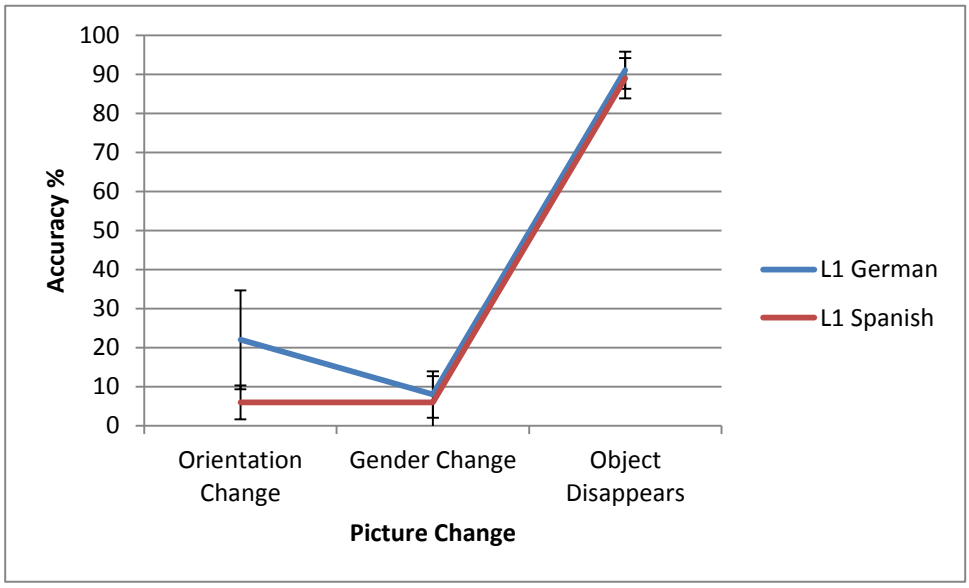
Table 5: Picture recognition task accuracy percentages and RTs for L1 and L2 speakers.

		% Accuracy Change in Object Orientation		% Accuracy Change in Facial Gender		% Accuracy Object Disappears		Average Reaction Time (ms)	
Group	**N**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
L1German	27	0.22	0.32	0.08	0.15	0.91	0.12	1668	501
L2GermanA2	58	0.38	0.29	0.14	0.19	0.85	0.22	1894	734
L2GermanB1	45	0.37	0.30	0.08	0.16	0.90	0.15	1833	678
L2GermanB2+	20	0.30	0.29	0.22	0.29	0.83	0.23	1769	617
L1 Spanish	27	0.06	0.11	0.06	0.17	0.89	0.13	1219	309
L2 SpanishA2	31	0.11	0.16	0.10	0.17	0.91	0.10	1583	537
L2 SpanishB1	52	0.15	0.30	0.12	0.25	0.95	0.07	1453	560
L2SpanishB2+	58	0.10	0.21	0.04	0.12	0.90	0.17	1338	498

L1 German and L1 Spanish

First, we examined accuracy scores. We ran a 2 (group: German, Spanish) x 3 (picture change: orientation, gender, object disappears) ANOVA on % of noticed picture changes (see Table 5). There was a significant main effect of group, $F(1,51)=447.377$, $p<.001$, $\eta_p^2 =.896$. On average, German (M=40%) and Spanish (M=34%) accuracy scores differed. There was a significant main effect for picture change $F(1,51)=1017.865$, $p<.001$, $\eta_p^2 =.951$. On average, accuracy differed with respect to orientation (M=14%), gender (M=7%) and object disappears (M=90%) changes. There was also a reliable interaction between picture change and group, $F(1,51)=3.855$, $p=.024$, $\eta_p^2 =.069$. Thus, German and Spanish speakers had different accuracy scores for different picture changes (see Figure 4). We made three comparisons to break down the interaction. Following the Bonferroni correction, the significance level was set at 0.05/3=.017. A T-test on % of noticed orientation changes for German (M=22 %) and Spanish (M=6%) speakers showed a significant difference, $t(1,51)=2.455$, $p=.017$, $d=.669$. A T-test on % of noticed gender changes for German (M=8%) and Spanish (M=6%) speakers showed no significant difference, $t(1,51)=.284$, $p=.77$, $d=.125$. A T-test on % of object disappears changes for German (M=91%) and Spanish (M=89%) also showed no significant difference, $t(1,51)=.440$, $p=.662$, $d=.153$.

Figure 4: Interaction between group (L1 German, L2 Spanish) and picture change (orientation, gender, object disappears). Error bars represent 95% confidence intervals.



Second, we examined RTs. Means and standard deviations are reported in Table 5. 19 (from 27) German speakers and 24 (from 27) Spanish speakers had missing RT data cause they had 0% accuracy scores for either orientation or gender or both. To analyze data in a full design with all participants, we collapsed RTs across orientation (90 responses), gender (44 responses) and delete object (588 responses) trials. We removed RTs further away than two standard deviations of each participants´ mean (21 responses; 2.9 % of total). We ran an independent samples T-test on the mean RT for correct responses. It showed there was a significant RT difference between German (*M*=1668; *SD*=501) and Spanish (*M*=1219, *SD*=309) speakers, $t(1,51)=3.971$, $p<0.001$, $d= 1.08$.

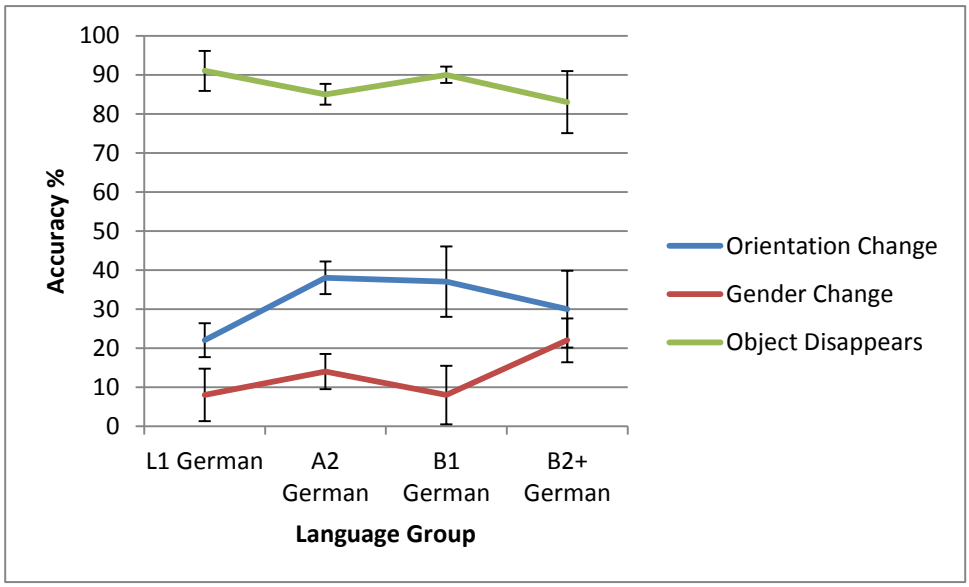
All in all, these data provide partial support for Hypothesis 1. L1 German speakers had a better recognition memory for object orientation than L1 Spanish speakers, but L1 Spanish speakers did not have a better recognition memory for facial gender of agents than L1 German speakers. On average, German RTs were higher than Spanish RTs, which means that longer looks to pictures by German participants may have contributed to the orientation effect.

L1 and L2 German comparison

First, we examined accuracy scores. We ran a 4 (group: L1, A2, B1, B2+) x 3 (picture change: orientation, gender, object disappears) ANOVA on the % of noticed picture changes. There was no significant main effect of group $F(3,146)=.087$, $p=.458$, $\eta_p^2 =.018$. Thus, on average, L1 speakers´ (*M*=40.3%) and A2 (*M*=45.7%), B1 (*M*=45%) and B2+ (*M*=45%) speakers´ accuracy scores did not

differ significantly. There was a significant main effect of picture change, $F(1,146)=899.371$, $p<.001$, $\eta_p^2=.860$. Thus, accuracy differed for orientation ($M=32\%$), gender ($M=13\%$) and object disappears ($M=87\%$) changes. There was also a reliable interaction between picture change and group, $F(3,146)=3.735$, $p=.013$, $\eta_p^2=.071$. Thus, L1 and L2 German speakers´ responses differed with respect to different picture changes (see Figure 5). We ran three One way ANOVAs comparing accuracy for orientation, gender and object disappears changes for L1 speakers and the three L2 learner groups (see Table 5 for percentages) to break down the interaction. A One way ANOVA on % of noticed orientation changes showed no significant difference with $F(3,149)=2.055$, $p=.109$, $\eta_p^2=.041$. A One way ANOVA on % of noticed gender changes showed a significant difference with $F(3,149)=2.895$, $p=.037$, $\eta_p^2=.056$. However, post hoc (Bonferroni) analyses revealed that differences between groups were not significant after all with $p>0.05$. A One way ANOVA on % of object disappears changes also showed no significant difference with $F(3,149)=1.418$, $p=.240$, $\eta_p^2=.028$.

Figure 5: Interaction between group (L1, A2, B1, B2+ German) and picture change (orientation, gender, object disappears). Error bars represent 95% confidence intervals.



Second, we considered RTs. Means and standard deviations are reported in Table 5. 68 (from 123) L2 German learners had missing RT data cause they had 0 % accuracy for either orientation or gender, or both. To analyze data in a full design with all participants we collapsed orientation (520 responses), gender (193 responses) and delete object (1306 responses) trials. We removed RT responses further away than two standard deviations of each participants´ mean (103 responses; 5,1% of total). We ran a one way ANOVA on the RT responses for the group of L1 speakers and the three groups of L2 speakers. It showed there was no significant RT difference between the different groups with, $F(3,146)=.747$, $p=.526$, $\eta_p^2=.015$.

Third, we also compared L1 and L2 German accuracy scores for the different sentence types: *legen* [lay] (prime1), *stellen* [stand] (prime2) and *es gibt* [there are] (neutral) (see Table 2 for full sentence examples). We distinguished between sentences with spatial language (*legen* [lay] and *stellen* [stand]) and sentences without spatial language (*es gibt* [there are]). We calculated accuracy percentages with respect to the number of trials (respectively eight and four trials) for these two sentence types. We ran a 4 (group: L1, A2, B1, B2+) x 2 (sentence type: spatial, not spatial) on % of noticed orientation changes. There was no main effect of group, $F(3,146)=2.392$, $p=.071$, $\eta_p^2=.047$. Thus, on average, accuracy for L1 speakers (M=21%) and A2 (M=38.1%), B1 (M=35.7%) and B2+ (M=28.5%) learners did not differ significantly. There was a main effect of sentence type, with $F(1,146)=13.357$, $p<.001$, $\eta_p^2=.084$. On average, accuracy was higher for sentences with spatial language ($M=35.8$, $SD=32.4$) as compared with sentences with no spatial language ($M=30.2$, $SD=30.3$) (see Figure 6). There was no significant interaction between group and sentence type, $F(3,146)=2.584$, $p=0.056$, $\eta_p^2=.05$. Thus L1 and L2 German speakers´ accuracy did not differ significantly for different sentence types (see Table 6 for mean accuracy per group and sentence type).

Figure 6: Main effect of sentence type (spatial, not spatial) for L1 and L2 German speakers. Error bars represent 95% confidence intervals.

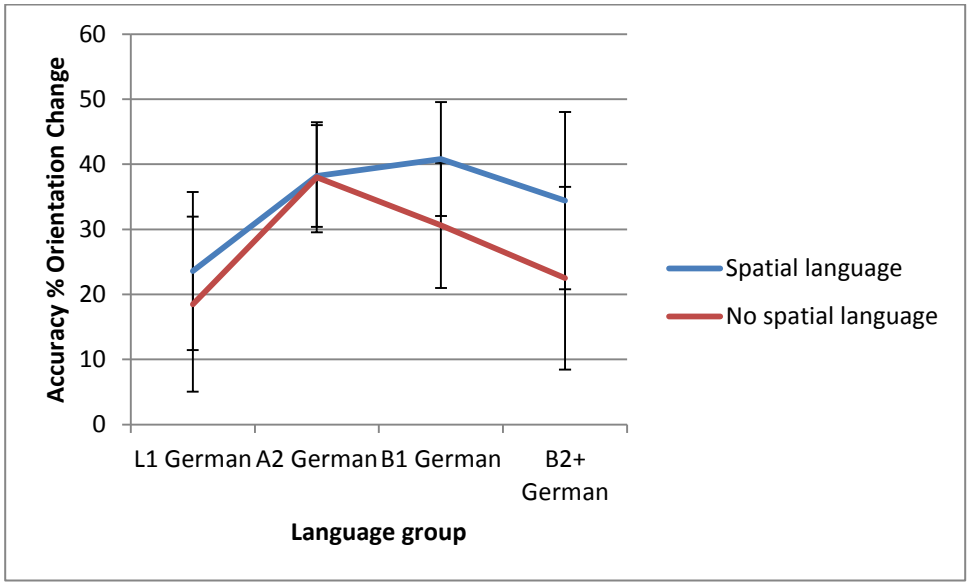


Table 6: Means and standard deviations for L1 and L2 German speakers for trials with and without spatial language

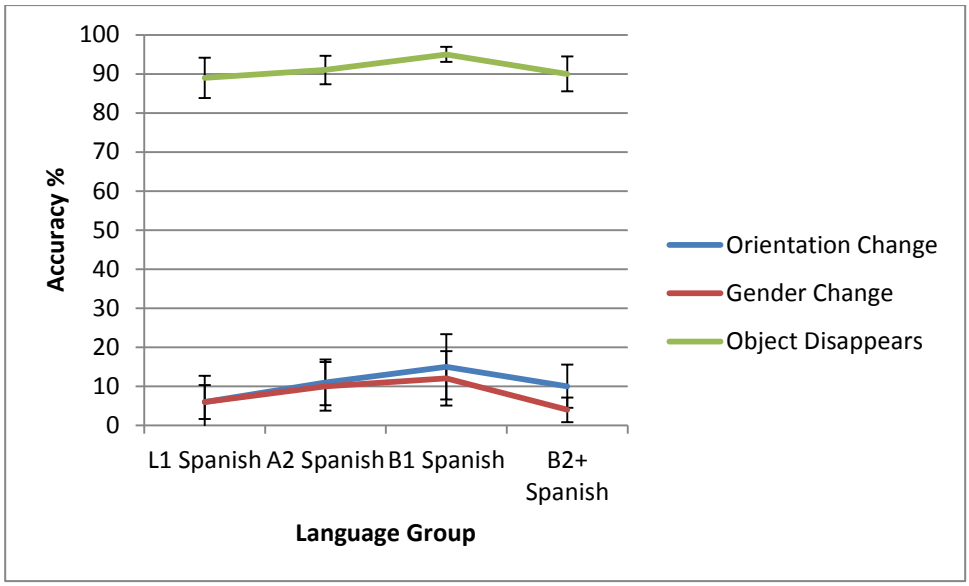
	N	Spatial language		No spatial language	
Group		**M**	**SD**	**M**	**SD**
L1 German	27	23.6	30.7	18.5	34.0
A2 German	58	38.2	29.7	38.0	32.1
B1 German	45	40.8	29.1	30.6	32.0
B2+ German	20	34.3	29.1	22.5	30.0

All in all, these data support Hypothesis 2, but not Hypothesis 3 and 4. Spanish L2 German learners´ memory was affected by object position marked by German verbs (Hypothesis 2). Analyses of RTs and sentence type provided further support for this notion. However, L2 learners´ memory was not worse than L1 German speakers´ memory (Hypothesis 3) and the memory effect was not mediated by L2 proficiency (Hypothesis 4). An ANOVA indicated that the different L1 and L2 proficiency groups did have different accuracy scores for the gender condition, but this was not confirmed by post hoc analyses.

L1 and L2 Spanish comparison

First, we considered accuracy scores. We ran a 4 (group: L1, A2, B1, B2+) x 3 (picture change: orientation, gender, object disappears) mixed ANOVA on the % of noticed picture changes. There was a main effect of group $F(1,167)=2.993$, $p=.033$, $\eta_p^2 =.052$. Thus, on average, L1 (M=34%), A2 (M=37%), B1 (M=41%) and B2 (M=35%) speakers´ accuracy differed (see Figure 7). However, post hoc analyses (Bonferroni) revealed no significant differences after all with $p>0.05$ for all comparisons. There was a main effect of picture change $F(1,167)=1121.716$, $p<.001$, $\eta_p^2 =.871$. Post hoc analyses (Bonferroni) revealed that accuracy scores for object disappears changes (M=91%) were significantly higher than those for orientation (M=11%) and gender (M=8%) changes, with p<.001 (see Figure 7). There was no significant interaction between group and picture change, with $F(1,167)=.314$, $p=.930$, $\eta_p^2 =.006$. Thus, L1 and L2 speakers´ responses did not differ significantly with respect to different picture changes.

Figure 7: Main effects of group (L1, A2, B1, B2+ Spanish) and picture change (orientation, gender, object disappears). Error bars represent 95% confidence intervals.



Second, we considered RTs. 105 (from 141) L2 German learners had missing RT data cause they had 0 % accuracy for either orientation or gender or both. To analyze the data in a full design with all participants, we collapsed RT son orientation (207 responses), gender (157 responses) and delete object (1561 responses) trials. We removed RT responses further away than two standard deviations of each participants´ mean (105 responses; 5,5 % of total). We ran a one way ANOVA on the RT responses for the group of L1 speakers and the three groups of L2 speakers. It showed there were significant RT differences between the different groups, with $F(3,163)=3.024$, $p=.031$, $\eta_p^2=.053$. Post hoc analyses (LSD) showed significant differences between L1 Spanish and A2 learners, with $p=.007$, 95% CI [-624.8969, -103.2367]; L1 Spanish and B1 learners, with $p=.051$, 95% CI [-469.0514, 1.0228] and A2 and B2+ learners, with $p=.03$, 95% CI [23.3121, 465.5536]. Means and standard deviations are reported in Table 5.

All in all, these data do not support Hypothesis 2, 3 and 4. German L2 Spanish learners´ memory was not affected by gender marked in Spanish (Hypothesis 2) and their memory was not worse than that of L1 Spanish speakers (Hypothesis 3). These data also do not support Hypothesis 4, as learners with different L2 proficiency all showed no recognition memory for changes in the facial gender of agents. This was the case despite longer looks (thus higher RTs) for A2 and B1 learners as compared with L1 Spanish speakers and longer looks for A2 learners as compared with B2+ learners.

5. Discussion

The two main theoretical questions in this study were: does bilinguals´ perception shifts with a change of language? And does L2 proficiency mediate the effect of the L2 on nonlinguistic cognition? We made use of cross linguistic variation in the expression of (motion in) space and gender in German and Spanish to investigate these questions. We examined L1 speakers and adult L2 learners with different levels of L2 proficiency. The cognitive process of relevance in this study was recognition memory. We investigated whether TFS effects could be found across L1 speakers and L2 learners in a fine-grained experimental task. We also investigated whether potential memory effects were mediated by L2 proficiency.

In the experimental task we registered in critical memory trials whether subjects recognized changes in pictures that were linked to linguistic cues given in trial phase 1. We predicted that L1 German speakers´ recognition memory of object orientation would be better than that of L1 Spanish speakers. For perception of facial gender of agents, we predicted this effect to be reversed (Hypothesis 1). Also, we predicted that L2 learners´ recognition memory would follow the spatial or gender aspects encoded in the L2 (Hypothesis 2); and that the effect would be more profound for L1 speakers as compared with L2 learners (Hypothesis 3). Finally, we predicted that L2 proficiency would mediate the effect of the L2 on recognition memory, so that L2 speakers with lower L2 proficiency would have worse recognition than L2 speakers with higher L2 proficiency (Hypothesis 4).

With respect to placement verbs and recognition of changes in object position, we found support for Hypothesis 1 and 2. We found that (a) L1 German speakers were significantly better to spot changes in object orientation than L1 Spanish speakers, whose score was close to 0% correct; and (b) Spanish learners of L2 German were as good as L1 German speakers to spot changes in object orientation. This effect was the same for learners at all levels of proficiency. Regarding the RT data, we found that L1 German RTs were significantly higher than L1 Spanish RTs. This means that longer looks at the pictures by L1 German speakers could have led to higher Germans scores. Thus we need to be careful in drawing conclusions. In the case of the L1 and L2 German speakers RTs were not significantly different. This means that we can be more confident in stating that spatial language was an important factor in the memory effect. Further support for this idea came from an analysis comparing trials with spatial language and trials with neutral language.

The orientation results comply with previous findings on motion verbs (Billmann et al., 2000, Gennari et al., 2002) in L1 speakers; and motion verbs in L2 speakers (Filipovic, 2011) and can be interpreted as evidence in favor of TFS. The results are contrary to those of Feist and Gentner (2007) who found that spatial prepositions lead English speakers to falsely recognize changed positions of objects in pictures; those of Coventry et al. (2010) who found no effects on recognition memory for Spanish-English linguistic instances describing relations of containment and support and Bosse and Papafragou (2010) that found no effects on memory for the German verbs *stehen* [sit] and *liegen* [lie]. The

difference between our findings and the results in these studies may be due to the dynamics implied by motion verbs as compared with static scenes. The dynamicity implied by motion verbs may increase motor activity in the cerebral cortex leading to heightened attention and recognition.

The finding that lower and higher proficient L2 learners´ recognition memory for object orientation was as good as that of L1 speaker was not in line with predictions of higher cognitive load in L2 speakers in general and at lower levels of L2 proficiency in specific and consequent effects on memory. Findings suggest that in this experiment, the cognitive load of language processing was equal for L1 speakers and lower and higher proficient L2 learners, leading to similar recognition memory accuracy. How can we reconcile our findings with research that indicates L2 processing differs at different levels of proficiency, with higher proficient learners approaching L1 speaker automaticity (Rossi, Gugler, Friederici & Hahne, 2006; Van Hell & Tokowicz, 2010; Foucart & Frenck-Mestre, 2013)? We may explain the discrepancy as follows. It is important to realize that in research on L2 processing participants are usually presented with sentences that contain syntactic or semantic anomalies. Consequently, studies have revealed differences in electrical brain activity to those anomalies. In our experiment, participants were presented with sentences that were grammatical correct and semantical unambiguous. Moreover, all L2 German learners received a linguistic instruction on the meaning of critical German verbs right before embarking on the memory experiment. Considering this, it may be argued that sentences and critical verbs were equally easy to process for both lower and higher proficient L2 learners. If this was the case, it is not surprising that we did not find the subsequent differences for recognition memory that we predicted between L1 and L2 speakers and lower and higher proficient L2 learners.

With respect to gendered personal pronouns and recognition of facial gender, Hypothesis 1 was only partially supported and Hypothesis 2, 3 and 4 were not supported. As expected, L1 German speakers had scores close to 0% correct (which supports Hypothesis 1), but L1 and L2 Spanish also showed scores close to 0% correct, which means they hardly recognized any changes in the facial gender of agents. This was the case despite significant RT differences, thus shorter and longer looks, between the L1 Spanish and L2 Spanish groups. This does not comply with effects reported by Boroditsky, Schmidt and Philipps (2003). However, their study focused on grammatical gender encoded in object nouns and not on gendered personal pronouns. Possibly, a higher saliency (determined by factors as frequency of appearance in input, word class, position within sentence etc.) of linguistic forms may have driven the effect in the study by Boroditsky, Schmidt and Philipps (2003) whereas lower saliency of linguistic forms in our study lead to no effects. Let us take frequency of occurrence as example. Berman and Slobin (1994) argue that the less frequent forms appear in the input the less salient they are. Spanish is a “pronoun-dropping” language, which means pronouns may be omitted from speech or writing unless they are needed for purposes of emphasis or contrast (Butt & Benjamin, 1998). Thus it is plausible that the Spanish object nouns used in the studies of Borodistky and colleagues appear more frequently in language input than the personal pronouns *ellos* [they-masculine] and *ellas* [they-feminine]. The

difference in salience of critical forms in the separate studies would explain the discrepancy in results.

There are at least two (further) explanations for the lack of an effect for gender, but positive results for spatial language. A plausible explanation concerning the L2 learners is that the instruction factor was not constant. L2 German learners received a linguistic instruction on German placement verbs, whereas L2 Spanish learners did not receive such an instruction on Spanish personal pronouns. This was done for a good reason, namely that the meaning of the Spanish *ellos*/*ellas* [they-masculine/-feminine] is typically taught at A1-level (CEFR) whereas the German verbs are not. Nonetheless, the discrepancy in instruction arguably resulted in a situation where the L2 German speakers' attention to the linguistic differences was greater than that of L2 Spanish speakers. The novelty of (the meaning of) the German verbs or reinforcement of already existing knowledge in L2 German learners may have caused the verbs to be more salient, causing the language effect. There was no novelty or reinforcement aspect for L2 Spanish learners, leading Spanish pronouns to be less salient than the German verbs, hence no effect. In future studies, instruction conditions should be kept equal across different learner groups to maintain the largest degree of comparability among the language dimensions investigated.

An alternative explanation concerning both L1 and L2 speakers is the following. There is both experimental and neurological evidence indicating that humans have a specific memory system dedicated to human faces that may qualitatively differ from other types of memory (Beyn & Knyazeva, 1962; Yin, 1969; Engstler & Engstler-Schooler, 1990). In fact, Engstler and Engstler-Schooler (1990) found that descriptions of faces worsened individuals´ memory for characteristics of those faces. Thus, a different nature of recognition memory for (the position of) objects and (gender characteristics of) faces may also explain the discrepancy between space and gender results.

Returning to our two main theoretical questions, we can state the following. The present study shows that the L2 indeed may affect bilingual´s perception and lead attention to aspects encoded in the L2. Consequently, this may affect recognition memory of aspects encoded in the L2. We found that L2 German learners´ recognition memory for object orientation was as good as L1 German speakers´ recognition memory. We found no such effects in relation to recognition of changes in facial gender of agents. Plausibly, this was due to differences in instruction or a different type of memory system. The presented data in this study does not support the notion that L2 proficiency mediates effect of the L2 on nonlinguistic cognition. L2 German and L2 Spanish learners with different L2 proficiency did not show significant differences in recognition memory accuracy.

6. Conclusion

This study has shown new empirical evidence that informs us about language-specific effects in L1 speakers and adult L2 learners. There are three main findings. First, not gender, but spatial language affects recognition memory for object position in L1 speakers and L2 speakers. The latter result can be interpreted as a TFS effect. Second, L2 German speakers´ recognition memory for object position is as

good as that of L1 German speakers and third, this effect was found not to be mediated by L2 proficiency. In future work, employing a similar recognition task, one could consider presenting L2 participants with complex sentences or longer pieces of text instead of simple sentences (as proposed for research on language comprehension in general by Zwaan, 2014). This may increase processing or cognitive load for learners with lower levels of L2 proficiency as compared with learners with higher levels of L2 proficiency. Alternatively, well in advance before participants complete the memory experiment, one may differentiate between learners who already know the form and those who don't and only instruct the latter group. In this case, it may be revealed that recognition effects are mediated by L2 proficiency. Future studies along such lines need to unravel and consolidate findings within this paper and the linguistic relativity literature in general.

Author contributions

All authors contributed to the study design. Testing and data collection were performed by Author 1. Author 1 and 3 performed data analysis. Author 1 drafted the manuscript and Author 2 and 3 provided critical revisions. All authors approved the final version of the manuscript for submission.

Acknowledgements

This work has been supported by the EU 7th Framework Programme Marie Curie Initial Training Networks grant Nr. 316748 under the project *Language and Perception*. We thank our collaborators, hosts and local assistants who enabled data collection at the various test locations. We thank David Perrett and his lab for morphing the gender stimuli.

Endnotes

1. In the pilot study 12 participants were asked to consider 112 object combinations. They were asked how they would place the first object in each combination onto the second object. They responded by ticking a box on a five point scale that ranged from 1, "lying" to 3, "either way" to 5, "standing". For purposes of the current study we selected four object combinations that received an average rating of 3, "either way".

References

Abutalebi, J., Cappa, S., Perani, D. (2005). What can functional neuroimaging tell us about the bilingual brain? In J. Kroll & A. de Groot (Eds.), *Handbook of bilingualism* (479-515). Oxford: Oxford University Press.

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., Riches, N. & Thierry, G. (2015). Two languages, two minds: flexible cognitive processing driven by language of operation. *Psychological Science*. 26, 4, 518-526.

Bernstein, D. (2010). *Essentials of psychology*. Fifth Edition. Wadsworth Publishing.

Berthele, R. (2012). On the use of PUT verbs by multilingual speakers of Romansh. In A. Kopecka & B. Narasimhan (eds.) *Events of putting and taking: a crosslinguistic perspective* (145-166). Amsterdam, Philadelphia: John Benjamins.

Beyn, E., Knyazeva, G. (1962). The problem of prosopagnosia. *The Journal of Neurology, Neurosurgery, and Psychiatry*, 25, 154-158.

Billman, D. Swilley, A., Krych, M. (2000). Path and manner priming: verb production and event recognition. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society.*

Buchler, N., Reder, L. (2007). Modeling age-related memory deficits: a two-parameter solution. *Psychology and Aging*, 22, 104-121.

Boroditsky, L., Schmidt, L., Philipps, W. (2003). Sex, syntax and semantics. In Gentner, D. & Goldin-Meadow, S. (Eds.). *Language in mind: Advances in the study of language and thought* (61-79). Cambridge: MA: MIT Press.

Bosse, S., Papafragou, A. (2010). Spatial position in Language and Visual Memory: A Cross-Linguistic Comparison. *Proceedings from the 32nd Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Cadierno, T., Ibarratxe-Antunano, I., Hijazo-Gascón, A. (2016). Semantic categorization of placement verbs in L1 and L2 Danish and Spanish. *Language Learning*, 66, 1, 191-223.

Cook, V., Bassetti, B. (Eds.) (2011). *Language and Bilingual Cognition*. New York: Psychology Press.

Coventry, K., Valdés, B., & Guijarro-Fuentes, P. (2010). Thinking for speaking and immediate memory for spatial relations. In Z.-H. Han & T. Cadierno (Eds.), *Linguistic Relativity in L2 Acquisition: Evidence of L1 Thinking for Speaking,* pp.84-101. Clevedon: Multilingual Matters.

De Bruine, L., Jones, B., Little, A., Boothroyd, L., Perrett, D., Penton-Voaks, I., Cooper, P., Penke, L., Feinberg, D., Tiddeman, B. (2006). Correlated preferences for facial masculinity and ideal or actual partner´s masculinity. *Proceedings of the Royal Society B*, 273, 1355-1360.

Dufour, R., Kroll, J. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition*, 23, 166-180.

Fagan, S. (1991). The Semantics of the Positional Predicates liegen/legen, sitzen/setzen, and

stehen/stellen. *Die Unterrichtspraxis / Teaching German*, Wiley: American Association of Teachers of German, 24, 136-145.

Fausey, C., Boroditsky. L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin and Review*, 18 (1), 150-157.

Feist, M, Gentner, D. (2007). Spatial language influences memory for spatial scenes. *Memory & Cognition*, 35(2), 283-296.

Filipović, L. (2011). Thinking and remembering in one or two languages: bilingual vs. monolingual lexicalization and memory for motion events. *International journal of bilingualism,* 15(4), 466-485.

Filipovic, L. (forthcoming) Speaking in L2 but thinking in L1: Language-specific effects on memory for causation events in English and Spanish. *International Journal of Bilingualism.*

Francis, W. (2005). Bilingual semantic and conceptual representation. In: Kroll, J. & De Groot, A. (Eds.), *Handbook of Bilingualism* (pp. 251-267). New York: Oxford University Press.

Foucart, A., Frenck-Mestre, C. (2013). Language Processing. In J. Herschensohn & M. Young-Scholten (Eds.) The Cambridge Handbook of Second Language Acquisition (394-416). Cambridge: Cambridge University Press.

Gennari, S, Sloman, S., Malt, B, Fitch, W. (2002). Motion events in language and cognition. *Cognition,* 83, 49-79.

Gentner, D., Golding-Meadow, S. (Eds.). (2003) *Language in mind: Advances in the study of language and thought*. Cambridge: MA: MIT Press.

Gullberg, M. (2009). Reconstructing verb meaning in a second language. How English speakers of L2 Dutch talk and gesture about placement. *Annual Review of Cognitive Linguistics*, 7, 222-245.

Gumperz, J., Levinson, S. (1996). *Rethinking linguistic relativity*. Cambridge: Cambridge University Press.

Han, Z., Cadierno, T. (2010). *Linguistic relativity in SLA: Thinking for speaking.* Bristol, Buffalo, Toronto: Multilingual Matters.

Heredia, R., McLaughlin, B. (1992). Bilingual memory revisited. In R. Harris (Ed.), *Cognitive processing in bilinguals* (91-103). North-Holland: Elsevier Science Publishers.

Hicks, J., Marsch, R. (2000). Toward specifying the attentional demands of recognition memory. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 2000, 26(6), 1483-1498.

Ibarretxe-Antunano, I. (2012). Placement and removal events in Basque and Spanish. In A. Kopecka & B. Narasimhan (Eds.) *Events of putting and taking: a crosslinguistic perspective* (145-166). Amsterdam, Philadelphia: John Benjamins.

Kopecka, A., Narasimhan, B. (Eds). (2012). *Events of Putting and Taking, a crosslinguistic perspective*. Amsterdam: Benjamins.

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

Kroll, J., Van Hell, J., Tokowicz, N., Green, D. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language & Cognition*, 13(3), 373/381.

Lambert, W., Havelka, J., Crosby, C. (1958). The influence of language acquisition contexts on bilingualism. *Journal of Abnormal and Social Psychology*, 56, 239-244.

Maigiste, E. (1982). Automaticity and interference in bilinguals. *Psychological Research*, 44, 29-42.

Maigiste, E., (1985). Development of intra- and interlingual interference in bilinguals. *Journal of Psycholinguistic Research*, 14, 137-154.

MacWhinney, B. (2005). A unified model of language acquisition. In: Kroll, J. & De Groot, J. (Eds.), *Handbook of Bilingualism. Psycholinguistic Approaches*. (pp. 49-67). Oxford University Press.

Naveh-Benjamin, M., Craik, F., Guez, J., Dori, H. (1998). Effects of divided attention on encoding and retrieval processes in human memory: Further support for an asymmetry. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 24, 1091-1104.

Pavlenko, A. (2014). *The bilingual mind. And what it tells us about language and thought*. Cambridge: Cambridge University Press.

Rossi, S., Gugler, M., Friederici, A., Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: evidence from event-related potentials. Cognitive Neuroscience, 18(12), 2030-2048.

Schooler, J., Engstler-Schooler, T. (1990). Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology*, 22, 36-71.

Slobin, D. (1996). From "thought and language" to "thinking for speaking." In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity. Studies in the social and cultural foundations of language, vol. 17* (70-96). Cambridge: Cambridge University Press.

Talmy, L. (1975). Semantics and syntax of motion. In: John Kimball (Ed.), *Syntax and semantics*, 4, 181-238, New York: Academic Press.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description, vol. 3: Grammatical categories and the lexicon* (36-149). Cambridge: Cambridge University Press.

Talmy, L. (2000). Toward a cognitive semantics: Typology and process in concept structuring (Vol. II), MA: MIT Press.

Thompson, A. (2015). Are your participants multilingual? The role of self-assessment in SLA research. *Language in Focus Journal*, 1(1), 51-65.

Trueswell, J. & Papafragou, A. (2010). Perceiving and remembering events crosslinguistically: evidence from dual-task paradigms. *Journal of Memory and Language,* 64, 64–82.

Van Bergen, G., Flecken, M. (2017). Putting things in new places: Linguistic experience modulates the predictive power of placement verb semantics. *Journal of Memory and Language,* 92, 26-42.

Van Hell, J., Tokowicz, N. (2010). Event-related brain potentials and second language learning: syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research*, 26(1), 43-74.

Weinreich, U. (1953). *Languages in contact.* The Hague: Mouton.

Whorf, B., Bissel Carroll, J. (1956). *Language, Thought and Reality. Selected Writing of Benjamin Lee Whorf*. MIT Press.

Wolff, B., Holmes, K. (2010). Linguistic relativity. *WIREs Cognitive Science,* 2(3), 253-265.

Yin, R. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81,141-145.

Zwaan, R. (2014). Embodiment and language comprehension: reframing the discussion. *Trends in Cognitive Sciences*.

Appendices

Appendix 1: instruction on the German verbs *legen/stellen*

Die Verben *stellen, legen, liegen* und *stehen*

Wussten Sie schon?
Auf Deutsch gibt es Verben, die die Position von Objekten oder Personen beschreiben. Diese Verben zeigen an ob das Objekt oder die Person sich in einer vertikalen oder horizontalen Position befindet. Vier Verben, die oft benutzt werden, sind:

1. stellen poner en una posición vertical
2. legen poner en una posición horizontal
3. liegen estár en una posición horizontal
4. stehen estár en una posición vertical

Falls man die Position oder Bewegung noch genauer beschreiben möchte, kann man im Deutschen noch *gerade-* oder *hin-* hinzufügen (zum Beispiel: *geradestellen, geradelegen; hinstellen, hinlegen*).

Konjugationen

	stellen	**stehen**	**legen**	**liegen**
Ich	stelle	stehe	lege	liege
Du	stellst	stehst	legst	liegst
Er/sie	stellt	steht	legt	liegt
Wir	stellen	stehen	legen	liegen
Ihr	stellt	steht	legt	liegt
Sie	stellen	stehen	legen	liegen

Beispiel

Ulrich: "Uf, dieses Buch ist sehr groß und schwer!"

Kathi: "Du Armer! Komm, dann **legen** wir das auf den Tisch!"

Übung

Geben Sie an, ob die folgenden Verben eine horizontale, vertikale oder neutrale Position bezeichnen. Zeichnen Sie bitte einen Kreis um die richtige Antwort:

1. beantragen horizontal / vertikal / neutral
2. fordern horizontal / vertikal / neutral
3. geradelegen horizontal / vertikal / neutral
4. liegen horizontal / vertikal / neutral
5. zuweisen horizontal / vertikal / neutral
6. legen horizontal / vertikal / neutral
7. ergeben horizontal / vertikal / neutral
8. stehen horizontal / vertikal / neutral
9. hintereinanderstellen horizontal / vertikal / neutral
10. geradestellen horizontal / vertikal / neutral
11. wegstellen horizontal / vertikal / neutral
12. weglegen horizontal / vertikal / neutral
13. platzieren horizontal / vertikal / neutral
14. stellen horizontal / vertikal / neutral
15. hinstellen horizontal / vertikal / neutral
16. hinlegen horizontal / vertikal / neutral
17. machen horizontal / vertikal / neutral
18. aufeinanderlegen horizontal / vertikal / neutral

Paper 3:

Dietha Koster, Teresa Cadierno, Kenny Coventry, Marco Chiarandini

Do orientation and size matter? Mental simulation of object properties in L1 and L2 readers**.**

‘

Do orientation and size matter? Mental simulation of object properties in L1 and L2 readers

Dietha Koster , Teresa Cadierno , Kenny Coventry, Marco Chiarandini

Abstract: How do second language (L2) learners comprehend sentences in their L2? There is evidence for two different approaches to language comprehension in native (L1) speakers: abstract, amodal symbols (Burgess & Lund, 1997) and "mental simulation" (Barsalou, 2008). Zwaan (2014) argues that the contributions of these different approaches to language comprehension vary as a function of the degree to which language is embedded in the environment. In this study we investigate if and how sentences cause L2 speakers in a foreign language context to make mental simulations of object orientation and size, in comparison with L1 speakers. In addition, we explore whether individual differences in L2 learners affect processing of meaning. The experimental task involved reading sentences and responding to pictures of objects. L2 learners received an instruction on language-specific L2 forms before embarking on the experimental task. Results reveal three findings. First, L1 speakers process meaning faster than L2 speakers. Second, we found support for the hypothesis that both L1 and L2 speakers simulate object size, but not orientation. Third, we find (partial) evidence that proficiency and individual test scores on knowledge of language-specific forms affected speed of meaning processing. These findings show that L2 learners too may rely on simulations of salient object properties such as size during language comprehension. However, speed of meaning processing is different in L1 and L2 speakers and may be affected by individual L2 learner factors.

Keywords:
mental simulation, L2 learning, individual differences, form-meaning connections, object recognition

Highlights:
- L1 speakers process meaning faster than L2 speakers
- Both L1 and L2 speakers simulate object size but not object orientation
- Knowledge of language-specific L2 forms affects speed of meaning processing in L2 learners

1. Introduction

Successful second language (L2) learners are able to comprehend written messages in their L2. Yet, how do they do it? Do they draw on the same mechanisms as in their first language (L1) or do they rely on different processes? Up until fifteen years ago, the mainstream view was that the human mind handles language as a computer does. This means it combines abstract, amodal and arbitrary symbols (i.e.) words by syntactic rules (e.g., Burgess & Lund, 1997; Chomsky, 1980; Fodor, 2000; Kintsch,

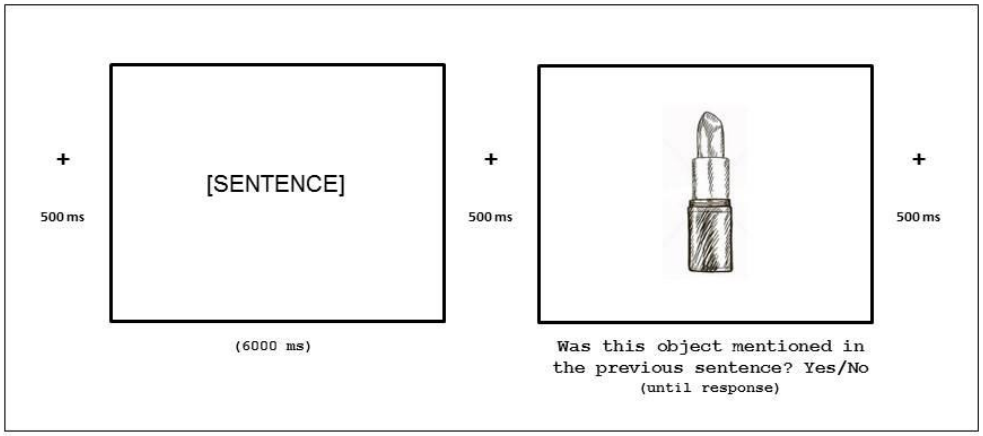
1988, Pinker, 1994). The main problem with this conceptualization of cognition is that it has no connection to actual experience. A classic example of this problem is the "Chinese Room" argument. Suppose a foreigner lands at a Chinese airport knowing none of the local language, but carrying a Chinese dictionary. When interpreting airport signs, the traveler will become stuck in an endless loop of abstract symbols, as every definition in his dictionary references to other symbols. This has been referred to as the "symbol grounding problem" (cf. Harnad, 1990). In recent years, "grounded cognition" theory has proposed another perspective on cognition. Following grounded cognition, human thought and language are shaped by our bodies and grounded in our perceptual experiences with the world (Barsalou, 1999, 2008; Glenberg, 1997; Glenberg & Robertson, 1999, 2000; Lakoff, 1987).

An important finding within grounded cognition is that so-called "mental simulations" may drive language comprehension (Zwaan & Pecher, 2012). For example, in perceiving a picnic, the human conceptual system might construe perceived individuals as instances of *tree, table, watermelon, eat, above* and so forth. To accomplish this, the conceptual system binds specific tokens in perception (i.e. individuals) to knowledge for general types of things in memory (i.e. concepts) (Barsalou, 1999). This process involves a reactivation of previous experiences with the world and is referred to as mental simulation. Empirical work has supported the idea that individuals make simulations of object orientation (Stanfield & Zwaan, 2001), shape (Zwaan, Stanfield & Yaxley, 2002; Engelen et al., 2011; Sato et al., 2013) and color (Zwaan & Pecher, 2012) when processing language. In these studies monolingual (L1) speakers read sentences like "The carpenter hits the nail into the floor". Subsequently they saw a picture of a nail that matched (vertical) or mismatched (horizontal) the object orientation implied by the context sketched in the sentence. The task was to determine whether the presented object was mentioned in the preceding sentences (see Figure 1). Results showed that participants had faster recognition of objects that matched the sentence as compared with mismatching instances (but see Connell, 2007 and Rommers et al., 2013 for different results). This suggests that L1 speakers may indeed reactivate previous knowledge of or experience with the world to understand language.

Zwaan (2014) argues that the contributions of abstract symbols and mental simulations to language comprehension vary as a function of the degree to which language is embedded in the environment. The more language is embedded in the environment, the larger the chance that mental simulation will be a route to comprehension. An unexplored question is to what degree adult L2 learners (such as the foreigner at the Chinese airport) rely on abstract, amodal symbols or grounded simulations during language comprehension. This question is particularly relevant in relation to cross linguistic differences, e.g., forms that have no direct translation equivalents. This question is also particularly relevant for L2 learners learning the L2 in a foreign language context. In such a classroom learning context, the learner may rely primarily on books as an L2 learning method and lack in worldly, perceptual experiences needed to understand the world described by the target language (Hymes, 1972; Collentine & Freed, 2004). This may lead the L2 learner to rely on abstract, amodal symbols instead of grounded simulations made by L1 speakers of the target language. Besides learning context, input

factors (e.g. frequency, saliency) and learner factors (e.g. proficiency, motivation) may affect comprehension (VanPatten, Williams & Rott, 2004). It is plausible that L2 learners with greater exposure to L2 input outside the classroom and with a higher degree of L2 proficiency and motivation to engage with the L2 may rely more on grounded simulations than lower proficient and less motivated learners, who are exposed to a lesser amount of L2 input.

Figure 1: Example of an experimental sentence-picture verification trial



In this paper we investigate if and how L1 and L2 speakers make mental simulations of object properties during sentence processing. First, we introduced L2 learners to L2 specific forms that describe object orientation and object size. Second, L1 speakers and L2 learners completed a behavioral task where they were presented with sentences and pictures of objects. The sentences contained information about object orientation and size that either matched or mismatched with an object present subsequently. The task was to determine whether the presented object was mentioned in the preceding sentences. See Figure 1 for an example of an experimental sentence-picture verification trial. We registered Yes or No answers and how fast participants recognized the object. We compared L2 results with L1 speaker results. Moreover, we explored whether and how individual differences in L2 speakers´ proficiency, knowledge, L2 use and motivation affected their reaction times (RTs). In the following section we discuss previous research relevant to our study, covering theoretical and empirical issues related to form-meaning connection making in L1 and L2; the role of L2 instruction; L1 and L2 processing; and work on mental simulation in L1 and L2. This review will lead us to present our hypotheses.

2. Theoretical background

2.1 Form-meaning connections in L1 and L2

Representation of meaning is central to language comprehension at all levels (e.g. word, sentence,

text). But how do L1 and L2 speakers represent meaning? Both L1 and L2 speakers need to establish form-meaning connections (FMCs). A FMC is made when an individual registers a linguistic form, a meaning and the fact that the form encodes the meaning in some way (VanPatten, Williams & Rott, 2004).

L1 and L2 acquisition differ in numerous aspects. A critical difference can be found in language input: The typical L1 pattern of acquisition results from naturalistic exposure in situation where caregivers naturally scaffold development (Tomasello & Brooks, 1999), whereas classroom environments for foreign language teaching can distort the patterns of exposure, function, medium and social interaction (Ellis & Laporte, 1997). VanPatten et al. (2004) claim that FMCs may be placed on any point on various continua: from partial to complete, from weak to robust and from non-target like to target like. While L1 speakers may develop complete, robust and target-like FMCs; L2 learners may create incomplete (Shirai & Andersen, 1995), less than robust (Baddeley, 1990) and non-target like (Bardovi-Harlig, 2000) connections. Within L2 learners, factors that may affect FMC during L2 acquisition may fall into two broad categories: input factors (e.g., frequency, saliency) and learner factors (e.g. proficiency, motivation) (Ellis, 2004). All in all, it is plausible that not all L1 and L2 speakers draw upon FMCs that are equally complete, target like and robust.

L2 Forms without direct L1 translation equivalents may arguably pose challenges for L2 learners. Let us take a Spanish sentence example such as *Harry pone el libro en el estante* (Harry puts the book on the shelf). In German, one is obliged to choose either the verb *legen* [lay] or *stellen* [stand](Fagan, 1991; Lemmens, 2006) to describe this event. Contrary to the Spanish verb *poner* (´put´), *legen* and *stellen* indicate horizontality or verticality of the placed object with respect to the referent object, the surface on which it is being placed. Thus a Spanish learner of L2 German would need to establish a FMC, containing information about object orientation, which is not present in the L1. Cadierno et al. (2016) have observed learning difficulties of placement verbs by Spanish learners of Danish (which also distinguishes between *lægge* [lay] and *sætte*[stand]. Now take the sentence *Harry puts a large book on the shelf*. In Spanish, one can express largeness by adding an augmentative suffix like *–ón*, *-azo* or *-ote* to masculine nouns, or *-ona-, -aza* or *–ota* to feminine nouns (Gooch, 1967), so one can have a *libro* (´book´) or a *libr<u>ote</u>* [book-large]. Spanish augmentative suffixes are mainly used to denote large size, but they can also add an emotional tone to a given word, mostly associated with either admiration or a pejorative idea of clumsiness, unpleasantness, awkwardness, excess etc. (Butt & Benjamin, 2005; Hualde et al. 2010). This type of morphological marking indicating large size is not possible in German (Lohde, 2006, Korecky-Kröll & Dressler, 2007). Thus, in this case, a German learner of L2 Spanish would need to forge a FMC, containing information about object size, which is not present in his or her L1.

Within the field of SLA a hotly debated issue has been whether L2 instruction is effective or not (Benati & VanPatten, 2010). However, in 2001, Norris and Ortega reviewed over 40 published studies

on the effects of instruction and concluded that overall, instruction is effective in helping L2 learners to establish L2 form-meaning connections. They note that instructional techniques that emphasize meaning are more commensurate with what psychologists know about how the brain internalizes new knowledge (Norris & Ortega, 2001). De Knop and Perrez (2014) developed a meaning-based L2 instruction for teaching of German posture verbs and found that the instruction was successful in facilitating the learning and adequate use of these verbs. Marcos Miguel (2010) found that L2 Spanish learners were graphically aware of augmentative morphology after receiving L2 instruction (type of instruction unknown). This shows that at least, L2 instruction may lead L2 learners to register augmentative suffixes.

Following theorizing on form-meaning connections and work on L2 instruction we may predict the following. First, we may predict that form-meaning mappings are stronger in L1 speakers than in L2 speakers. Moreover, individual differences in L2 learners may lead to variation in whether mappings are complete, targetlike and robust. It is plausible that cross linguistic differences between L1 and L2, as the ones described above, provide a particular challenge for L2 learners. Research suggests that L2 instruction may be effective in helping L2 learners establish relevant form-meaning mappings.

2.2 L1 and L2 Processing

Another relevant question for our study is how L1 and L2 speakers process language. An important line of research investigates whether processing an L1 and an L2 involves separate processors or a single processor (or a combination of both). Most studies support the single processor claim. For example, Maigiste (1982, 1985) found that L2 speakers took more time than L1 speakers to process verbal material in each language. Soares and Grosjean (1984) found support that L2 speakers search both L1 and L2 lexicons in L1 processing. In addition, studies have shown that semantic comparisons between words from different languages took no longer than comparisons between words of the same language (Caramazza & Brones, 1980; Dufour & Kroll, 1995; Potter et al., 1984). Abutalebi, Cappa and Perani (2005) found neurological evidence that bilinguals with equal practice in both languages from birth, processed L1 and L2 through a single processing system.

Several empirical studies indicate that L2 processing changes during acquisition in late or adult L2 learners. It has been found that in early stages of L2 learning, lexical items of the L2 are processed through association with their translation equivalents in the L1, whereas in later learning stages (and with increased proficiency), processing of L2 words is more directly conceptually mediated (Dufour & Kroll, 1995; Kroll & Stewart, 1994). Thus, L1 and L2 lexical items are both thought to access a common semantic system directly as a L2 learner becomes more proficient in the L2. This notion has been expressed in Kroll et al.´s (2010) Revised Hierarchical Model (RHM) on L2 processing. There is also neurological evidence for L2 proficiency effects. Abutalebi, Cappa and Perani (2005) summarize neurological studies into L2 processing and conclude that proficiency is determinant in the cerebral representation of languages. Work on sentence processing has shown qualitative and quantitative

differences in processing of sentences between lower and higher proficient learners (for reviews see Van Hell & Tokowicz, 2010; Foucart & Frenck-Mestre, 2013).

Previous research on L1 and L2 processing leads to the following predictions. First, we may predict that L1 and L2 speakers process language differently. In particular, as L2 speakers may search both lexicons, they will process language *slower* than L1 speakers. Second, we may predict that L2 speakers with different levels of proficiency process language differently. In particular, speakers with lower levels of L2 proficiency may process the L2 through translation equivalents in the L1 whereas speakers with higher levels of L2 proficiency may process the L2 through a common semantic system, shared with the L1. Therefore lower proficient learners might process the L2 *slower* than higher proficient learners.

2.3 Simulation in L1 and L2

Does language cause mental simulation in L1 and L2 speakers? Stanfield and Zwaan (2001) developed a clever experimental task to provide empirical support for mental simulation (see Figure 1). Participants read sentences where the orientation of an object was implied by the context. For example, in *Harry puts the book on the shelf*, the orientation of the book is not specified, but world knowledge would lead one to think the book placed in the shelf upright. Following such sentences participants were presented with a picture (e.g., of a book), with the picture presented either horizontally or vertically. The task to respond (by pressing buttons) to respond to the question "Was this object mentioned in the preceding sentence?" Stanfield and Zwaan (2001) found that participants were faster to respond to this question if the orientation of the picture matched the implied orientation of the object in the sentence, supporting the notion that participants had mentally simulated object orientation.

It is important to recognize that different components of an utterance may drive simulation. Research on L1 speakers has distinguished a critical role for lexical items (nouns and verbs) and sentential context. To begin with words (nouns, verbs), it is known that individual lexical representations become active during language processing. Activation spreads through associations with perceptual and motor representations of modality-specific experiences they refer to, possibly through Hebbian learning (Colunga & Smith, 2005, Pülvermuller, 2012). Sato et al. (2013) found that Japanese motion verbs affected mental representations of object shape during sentence processing in L1 speakers. Considering sentential context, a number of authors have theorized a prominent role for grammar (Kaschak & Glenberg, 2000; Feldman, 2006) and Bergen and Wheeler (2010) found that grammatical aspect affected mental simulation.

There is little work on mental simulation in bilinguals, let alone on bilinguals with different levels of proficiency. Vukovic and Williams (2014) found with the standard sentence-picture verification task that advanced Dutch speakers of L2 English activated task-irrelevant meanings of interlingual homophones. For example, after hearing the English sentence "On the plate in front of you/at the far end of the table, you can see a bone", participants would see a picture showing a bean – the word for

which in Dutch is “boon” /bo:n/. The subsequent picture depicted a bean that varied in size (large/small), such that it mis/matched the distance implied by the different sentence introductions. Results showed that participants were slower to reject critical items where perceptual features matched the implied distance relationship. This indicates L1 mental simulations may take place during L2 processing. This is in line with L2 research indicating that possible words from different languages temporarily become active during reading and that access to L2 words is non-selective and automatic (i.e. not under the control of the reader) (Dijkstra, 2005).

Tomzcak and Ewert (2015) studied whether L1 speakers of Polish and English and advanced Polish learners of L2 English simulated real versus fictive motion. An example of real motion is “John runs through the forest”, an example of fictive motion is “The road runs through the forest” (Talmy, 2000). They predicted that processing of fictive motion involves simulation of physical motion and would therefore take longer than processing of static pictures and real motion for both L1 and L2 speakers. They presented participants with prime words (e.g. a verb indicating horizontal or vertical motion of the Figure) that matched with a following Polish or English sentence respectively. They asked participants to make meaning judgments about these sentences in Polish or English and registered Yes and No answers and RTs. They found that both L1 and L2 speakers had longer RTs for fictive motion trials as compared with static pictures and real motion. They interpreted this result in favor of mental simulation of motion in both L1 and L2 speakers.

The studies by Vukovic and Williams (2014) and Tomczak and Ewert (2015) suggest that proficient bilinguals may make L1 and L2 simulations during L2 comprehension. Vukovic and Williams (2015) have also shown that individual differences in L1 speakers may lead to different simulation patterns. They identified preferential usage of egocentric and allocentric reference frames in individuals and found in one of their experiments that only the egocentric group showed a match effect on the standard sentence-picture verification task. The allocentric participants did not produce evidence of the same effect. This indicates that individual differences may affect whether individuals make mental simulations. The findings in these three studies lead to wonder whether individual differences in L2 speakers can affect if and how they make mental simulations.

In this study we present L1 and L2 German speakers with sentences containing *legen/stellen* and L1 and L2 Spanish speakers with sentences containing augmentatives in a sentence-picture verification task as in Stanfield and Zwaan (2001). Following theory just reviewed, our main hypotheses are the following:

1. L1 and L2 German speakers will simulate object orientation whereas L1 and L2 Spanish speakers will simulate object size.

2. Individual differences in L2 input and L2 learner factors will affect how fast L2 learners process meaning.

Also, we predict that :

3. L1 speakers will process meaning faster than L2 speakers.

3. The current study

In this section we present information on two experiments that we designed. In Experiment 1 we investigate if and how L1 German speakers and Spanish learners of L2 German simulate object orientation. In Experiment 2 we explore if and how L1 Spanish speakers and German learners of L2 Spanish simulate object size. All participants were presented with a sentence-picture verification task (as in Stanfield & Zwaan, 2001). L2 learners participated in a learning activity on the meaning of German placement verbs and Spanish augmentative suffixes before embarking on the experimental task.

Participants
In Experiment 1, L1 German speakers (mean age: 24.0) and 122 Spanish students of L2 German participated. The L1 speakers were recruited at the University of Bremen in Germany (none knew Spanish); the L2 speakers studied German at the University of Seville or Granada in Spain. In Experiment 2, 34 L1 Spanish speakers (mean age: 21.2) and 100 German students of L2 Spanish participated. The L1 speakers were students at the University of Seville in Spain (none knew German); the L2 speakers studied Spanish at University of Münster or Humboldt University Berlin in Germany. We paid subjects a nominal fee for participation. Details of L2 participants are given in Table 1.

Table 1: Summary of background information for L2 participants in Experiment 1 (N=122) and 2 (N=100). L2 Motivation was self-rated on a scale from 1-5, with higher ratings indicating higher motivation.

	Experiment 1: L2 German				**Experiment 2: L2 Spanish**			
Participant Factor	**Min**	**Max**	**Mean**	**St. dev**	**Min**	**Max**	**Mean**	**St. dev.**
1. Age	18	35	21.3	4.0	18	57	24.3	5.1
2. Length of L2 learning (years)	1	16	3.0	2.0	0	22	4.4	3.5
3. Daily L2 Use (hours)	0	11.5	1.6	2.1	0	4	0.8	0.9
4. L2 Motivation	1	5	3.9	0.7	2	5	4.2	0.7
5. Pretest Score (% correct)	11.1	100	60.0	21.3	17.0	100	78.4	22.5
6. Posttest Score (% correct)	22.2	100	90.8	17.2	44.0	100	96.7	8.1

Materials

Experimental items consisted of sentences, followed by black-and-white drawings of objects. The essential aspect of each experimental trial was to answer 'yes' or 'no' to the question as to whether a preceding sentence had mentioned the drawn object (see Figure 1). Eight critical objects were identified in a pilot study as orientation-free objects (e.g. lipstick, battery, flashlight, bell, spool, deodorant, tube, glue stick). In the pilot study, participants indicated no preferred horizontal or vertical placement position when placing these object onto a Ground object (e.g. a plate).[1] This was to ensure that German speakers were only primed by the critical verbs *legen/stellen* [lay/stand] and not by a preferred or typical placement position of an object. In addition, the critical objects could appear in prototypical (3.5x3.5 inch) as well as in large (screen filling) size on a (17 inch) computer screen. In the black-and-white drawings, the 8 critical objects were presented in four different conditions: horizontal, vertical, large sized and small sized.

We then created critical sentences describing so-called ´placement events´ (e.g. *Mary puts the lipstick on the table*). In German, these sentences marked object orientation by either *legen* [lay] or *stellen* [stand], while Spanish sentences marked object size by including either an object noun with an augmentative suffix or without a suffix. See Table 2 for examples of critical sentence-picture pairs. In addition, we created filler sentences followed by black-and-white drawings of objects in their prototypical position. The filler sentences described people performing an action with an object that did not indicate object orientation or size. The filler objects were different from the 8 selected critical objects.

Table 2: Examples of critical sentence-picture pairs

Experiment	**Language**	**Object Property**	**Example Sentence**	**Picture Match**	**Picture Mismatch**
1	German	Orientation	Anna stellt die Lippenstift auf das Schneidebrett. [Anna stands the lipstick on the cutting board]		
2	Spanish	Size	Anna pone una barrota de labios en una tabla de cortar. [Anna puts the (big)lipstick on the cutting board.]		

Design

We aim to discover whether linguistic information that either matches or mismatches affects recognition time of subsequent pictured objects. For each object we created 4 sentences in German obtained by 2 (match/mismatch) x 2 (object orientation) and 4 sentences in Spanish obtained by 2 (match/mismatch) x 2 (object size). Thus, for each language we had 32 (4 x 8 objects) critical trials. To reduce the duration of the experiment, we distributed these trials over two lists (list A and B) with 16 trials each. In each list, we put 2 sentences per object by selecting the 2 matching or the 2 mismatching sentences for that object. Since the pilot study showed that objects were neutral with respect to orientation, we will not consider objects as a factor in the analysis.

The critical 16 trials per list were 'Yes' sentences on the 8 critical objects (we expected participants to answer ´Yes´, since the mentioned object was being depicted). We augmented each list with 16 further 'Yes' trials for the same 8 objects, yielding 32 trials for the 8 objects. We further augmented the lists with 32 'No' trials for the same 8 critical objects (here, an object appeared that was different from the one previously mentioned, thus the expected answer was ´No´). Finally, we introduced 64 further trials per list for 8 different objects. Again, on half of the trials we expected ´Yes´ answers; on the other half ´No´ answers. In total, each list contained 128 trials that were presented to the subjects. Table 3 shows the distribution of subjects among the lists.

Table 3: Distribution of L1 and L2 speaking participants among two lists with experimental trials (A and B)

Experiment 1: Orientation			**Experiment 2: Size**		
Language	**List**	**N**	**Language**	**List**	**N**
L1	A	13	L1	A	18
	B	17		B	16
L2	A	74	L2	A	47
	B	48		B	53

General Procedures (L1 and L2)

Participants took part in the computer experiment in groups of 5-20 in a quiet computer room. L1 speakers were instructed in their L1 and L2 learners in their L2. The instructions were: ´In this experiment you will read a number of sentences, followed by pictures of objects. Your task is to determine whether the shown object was mentioned in the sentence you read before. Please answer the question ´Was this object mentioned in the previous sentence?´ Give your answer by pressing Q(Yes)

or P(No).´ Participants were told that RTs were being measured; they were asked to make decisions about the pictures as quickly as possible and to keep their fingers on the Q and P button during the whole experiment.

L2 Procedures

L2 learners participated in two instructional activities before embarking on the computer experiment. This was to ensure that all L2 learners had explicit knowledge about the target forms and their meanings. First, they received a class-fronted instruction led by the experimenter in the L2, where the relevant forms (posture verbs or noun suffixes) and their meanings were discussed (Appendix 1). Second, learners studied a randomized list with 30 L2 words (15 target words) with L1 translations (Appendix 2). We examined L2 learners´ knowledge of the target forms by means of an 18-item pretest administered before instruction and a post-test (with items re-ordered) administered after class instruction. In the pre- and posttest, L2 German students indicated whether a given German verb (6 versions of *legen* [lay], 6 versions of *stellen* [stand] and 6 other verbs) indicated horizontality, verticality or no orientation. L2 Spanish students indicated whether a given Spanish noun (6 diminutive, 6 normal, 6 augmentative nouns) indicated small, normal or large size. After the instructional activities, the students completed the computer experiment as described above.

L2 Individual differences

After the computer experiment, L2 participants filled out a questionnaire (adapted from Gullberg & Indefrey, 2003, see Supplementary Material 1) where they reported background information on proficiency, language use, motivation and previous knowledge of the target forms. We selected one single measure for each of these constructs to use in subsequent analyses. The selection was made based on specificity (the more specific, the better) and objectivity (preference for test score over self-reported measure). This meant that proficiency was operationalized by the number of years a person had been learning the L2 inside and outside of class (other proficiency measures we had were self-reported level in terms of the Common European Framework of Reference (CEFR) and self-rated proficiency on a scale from 1-5. Exploratory analyses with these factors showed they did not reliably affect RT). Language use was measured in terms of hours of L2 use per day (speaking, hearing, reading and writing). L2 Motivation was based on the rating of three motivation statements on a five point scale (e.g. ´I like speaking [L2]´; ´I feel secure using [L2]´; ´I think it is important to know [L2]´). Knowledge of the target forms was operationalized through students´ pre- and posttest scores.

4. Results

We use linear mixed models to analyze RTs (measured in milliseconds) on critical orientation and size trials. We first analyze RTs for L1 and L2 German speakers for orientation trials (Experiment 1). Second we analyze RTs for L1 and L2 Spanish speakers for size trials (Experiment 2). Data were trimmed using the same procedures as in Zwaan and Pecher (2012), which means we only analyzed

Yes responses. See Table 4 for the mean RTs on critical trials for all groups. We make all analyses including the R script available as Supplementary Material.

Table 4: Mean reaction times (measured in milliseconds) and standard deviations on Match and Mismatch trials for L1 and L2 speakers

Group	**Match Mean**	**Std. dev**	**Mismatch Mean**	**Std. dev**
L1 German	705	335	764	438
L2 German	1092	348	1075	391
L1 Spanish	832	148	1321	147
L2 Spanish	996	573	1169	421

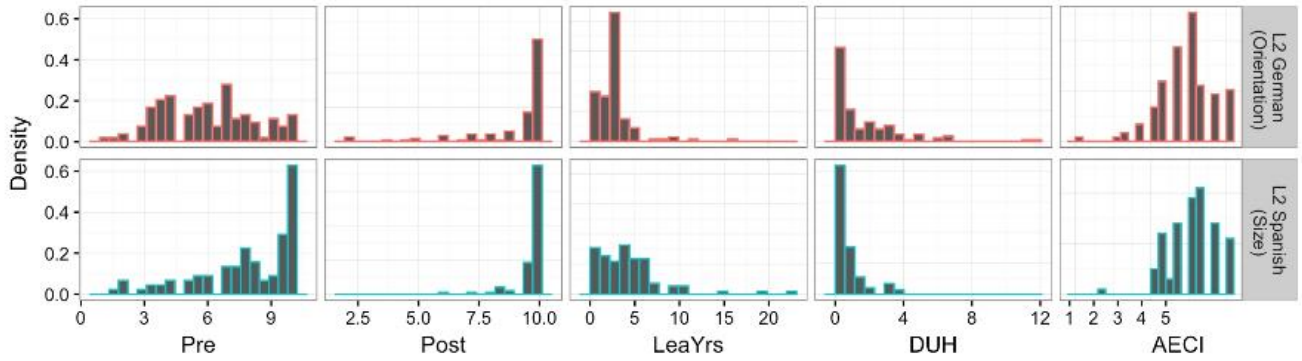
The distribution of RTs is right skewed and never reaches zero. Thus, we performed logarithmic transformation on the RT data to meet the assumptions underlying linear mixed models. For both Experiment 1 (orientation) and 2 (size) we fitted a model to study the differences between L1 and L2 speakers and one to investigate whether L2 individual factors affect performance. For each recorded data point, we considered a factor "Trial" to identify the specific sentence-drawing pair; a factor "Subject" to identify the subject to which the trial was administered; and a factor "List" to identify the two lists of trials that were presented to separate subjects. Further, we consider a factor "Type" to distinguish between match and mismatch trials and "Lang" to identify L1 and L2 speakers.

In the analyses on individual differences in L2 learners, we considered the following variables:

- "LeaYrs", the number of years spent learning the L2, as a proxy for L2 proficiency.
- "Pre" and "Post", the test scores in percentage correct on the pre- and posttest on knowledge of the target forms.
- "DUH", the number of hours of L2 use per day (daily usage hours).
- "AECI", the average rate of the perceived enjoyment, importance of and confidence in the L2, as a proxy for motivation.

The distribution of values for these variables in our data is shown in Figure 2. We rescaled the values of Pre and Post to be within 0 and 10, to make the values comparable with those of the other factors. Moreover, given the lack of knowledge a priori on how these factors interacted, we included all second order interactions in the analysis.

Figure 2: Distribution of individual difference factors for L2 learners (Experiment 1 and 2)



We specified our mixed effects models with random effects associated to “Trial” and “Subject” nested within "List:Type" and "List", respectively; and fixed effects associated with the other factors. We fitted our models in R using the package lme4 (R core team, 2015). We solved the model selection problem with the “dredge” function from the package MuMIn (Bartón, 2016). This function calculates the fit of a set of candidate models obtained by recombining in all possible ways the terms of a baseline model. It uses the Akaike Information Criterion (AICc) corrected for finite sample sizes to rank the exhaustive list of models and we finally selected the model that minimized the AICc value. For the models comparing L1 and L2 speakers it was possible to include all second order terms when specifying the baseline model. For the L2 individual factor models, computational issues excluded the possibility to include all second order terms. Therefore, we used knowledge on the experimental design to include only relevant random effects and a step backward procedure with likelihood ratio (REML) tests to discard fixed effects of second order until we reached an initial model of approachable size for exhaustive search.

We analyzed the final models that ranked first in the AICc criterion by means of t tests via the Satterthwaite approximation implemented by the R package lmerTest (Kuznetsova et al., 2016). We tested the random effects by means of the likelihood ratio test and they resulted always significant. We focus our discussion on the other terms that result statistically significant in these models (but see Bates et al., 2015).

Experiment 1: Orientation in L1 and L2 German

Comparing L1 and L2 German: Our dependent variable is reaction time (RT). After trimming (18.3% of) RT data for orientation mis/match trials we have a total of 1986 observations left. We note that for mismatch trials, the percentage of No responses for L1 German speakers was 3.3%; for L2 German speakers it was 18.1 %. The selected model (see Table 5) showed that the two random effects were

significant. The fixed effect Lang is significant (*Estimate*=-.347, *SE*=.056, *t*=-6.16, *p*<.001). L1 speakers reacted faster to both Match and Mismatch trials than L2 speakers. The fixed effect Type was discarded from the model already in the model selection phase. Thus responses to match or mismatch trials were not significantly different for both L1 and L2 German speakers.

Table 5: The selected model for L1 and L2 German data (Experiment 1)

```
## Random effects:
##  Groups          Name        Variance Std.Dev.
##  List:Subject    (Intercept) 0.0648   0.255
##  List:Type:Trial (Intercept) 0.0231   0.152
##  Residual                    0.1537   0.392
## Number of obs: 1993, groups:  List:Subject, 152; List:Type:Trial, 32
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)   6.8546     0.0371  80.5000  184.87  < 2e-16 ***
## LangL1       -0.3467     0.0563 142.7000   -6.16  6.9e-09 ***
```

Individual differences in L2 German speakers: We considered 1476 observations after trimming (24.4 % of the data was trimmed). In the selected model (see Table 6) the random effects are significant. For the fixed effects, we observe that LeaYrs (proficiency) (*Estimate*=-.053, *SE*=.015, *t*=-3.60, *p*=.0004), posttest (*Estimate*= -0.128, *SE*=.035, *t*=-3.60, *p*=.0004) and the interaction between pre- and posttest score (*Estimate*=.023, *SE*=.008, *t*=2.91, *p*=.0004) were significant. See Supplementary Material for extensive results.

Table 6: The selected model for L2 German data (Experiment 1).

```
## Random effects:
##  Groups          Name        Variance Std.Dev.
##  List:Subject    (Intercept) 0.0516   0.227
##  List:Type:Trial (Intercept) 0.0181   0.135
##  Residual                    0.1708   0.413
## Number of obs: 1476, groups:  List:Subject, 118; List:Type:Trial, 32
##
## Fixed effects:
##                     Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)          7.47978    0.47884 128.90000   15.62  < 2e-16 ***
## AECI                 0.19247    0.09739 118.70000    1.98  0.05045 .
## LeaYrs              -0.05309    0.01476 109.70000   -3.60  0.00049 ***
## List1               -0.32544    0.11423 110.90000   -2.85  0.00523 **
## Post                -0.12761    0.03545 123.00000   -3.60  0.00046 ***
## Pre                 -0.11088    0.09314 132.60000   -1.19  0.23601
## TypeMISMATCH        -0.14287    0.07293  29.70000   -1.96  0.05958 .
## AECI:Pre            -0.02424    0.01589 117.40000   -1.53  0.12994
## LeaYrs:List1         0.04568    0.02441 112.00000    1.87  0.06393 .
## List1:TypeMISMATCH   0.28358    0.10540  32.40000    2.69  0.01119 *
## Post:Pre             0.02279    0.00783 135.10000    2.91  0.00424 **
```

In Figure 3 we show the main effect of "LeaYrs" and in Figure 4 the interaction between pre- and posttest scores, which may explain the significance of the posttest score as well. We transformed RTs back into linear scale and added .95-confidence level bands. Figure 3 shows that RT decreases (thus reactions are faster) as proficiency (LeaYrs) increases. Considering knowledge of the target forms in Figure 4, we observe that students who scored low in both pre- and posttest (indicating they did not succeed in comprehending the L2 instruction) had the highest RTs (thus were the slowest). However, the RT of subjects with low pretest, but high posttest values is significantly lower (thus they reacted faster). The RT of subjects with high pre-test scores does not seem to be significantly affected by posttest score.

Figure 3: the effect of LeaYrs for L2 German (Experiment 1)

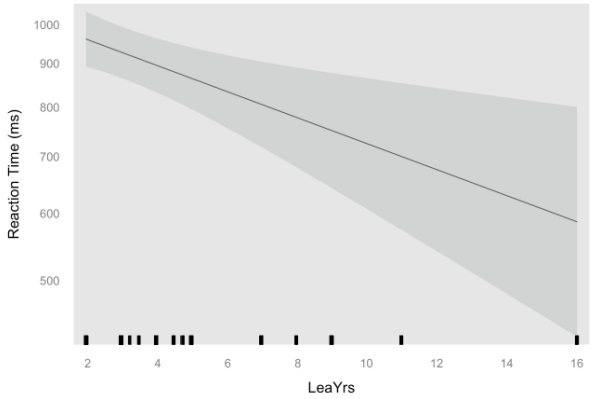
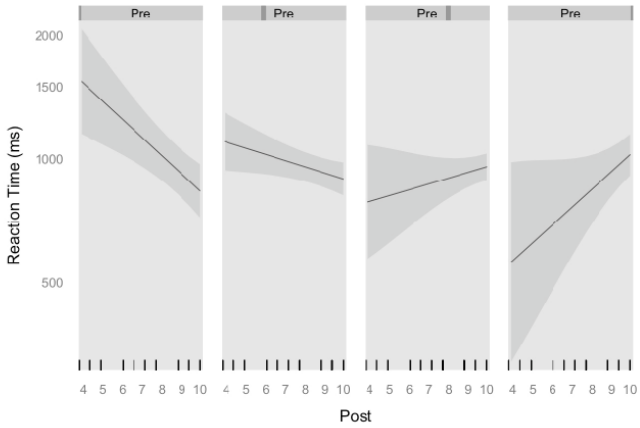


Figure 4: the interaction between pre- and posttest scores for L2 German (Experiment 1). Data are split up by the factor "Pretest" and facets are created for different levels represented by a vertical line in the strip text.



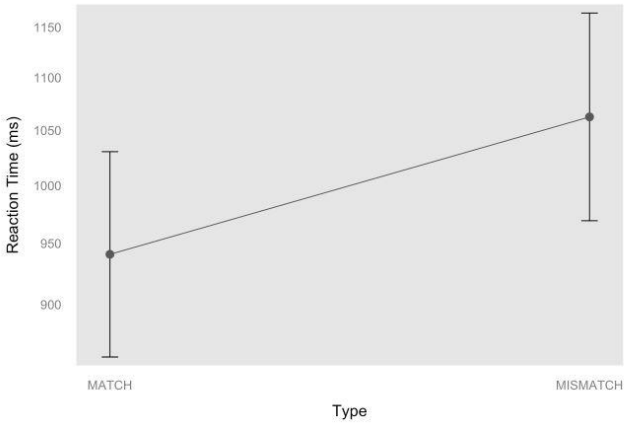
Experiment 2: Size in L1 and L2 Spanish

Comparing L1 and L2 Spanish: After trimming we have 1629 observations left (24.0 % of the data was trimmed). We notice that for mismatch trials, L1 Spanish speakers´ percentage of No responses was 18.9%, for L2 Spanish speakers it was 28,6%. The selected model (see Table 7) included the two random effects. The effect of Lang (*Estimate*=-0.2812, *SD*=0.0601, *t*=-4.7, *p*= .000) was statistically significant. L1 speakers had shorter RTs than L2 speakers on both Match and Mismatch trials. Here, there was also a significant effect of Type (*Estimate*=0.1216, *SD*=0.0560, *t*=2.17, *p*=0.037): L1 and L2 speakers reacted faster to Match trials (for L1 *M*=832; for L2 *M*=996) than Mismatch trials (for L1 *M*=1321; for L2 *M*=1169). Figure 5 shows the effect for Type. Also, Type was included in the five best models according to the AIC criterion (see Supplementary material).

Table 7: The selected model for L1 and L2 Spanish data (Experiment 2).

```
## Random effects:
##  Groups          Name        Variance Std.Dev.
##  List:Subject    (Intercept) 0.0802   0.283
##  List:Type:Trial (Intercept) 0.0221   0.149
##  Residual                    0.1394   0.373
## Number of obs: 1629, groups:  List:Subject, 134; List:Type:Trial, 32
##
## Fixed effects:
##              Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)    6.9257     0.0488  67.4000  141.92  < 2e-16 ***
## LangL1        -0.2811     0.0601 129.7000   -4.68  7.1e-06 ***
## TypeMISMATCH   0.1216     0.0560  31.6000    2.17    0.037 *
```

Figure 5: the effect of Type for L1 and L2 Spanish (Experiment 2). Data are split up by the factor "Type" and reported in different facets.

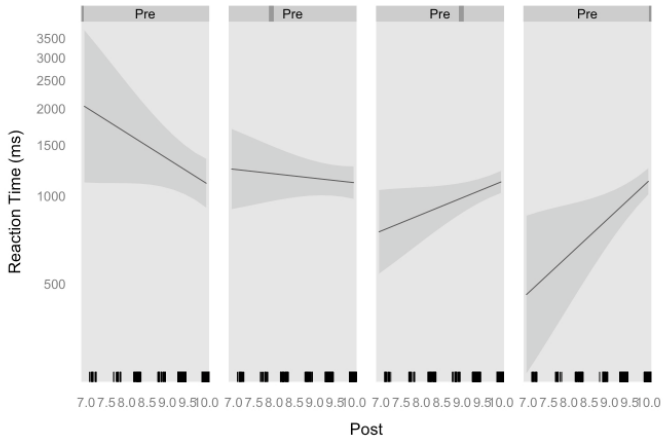


Individual differences in L2 Spanish speakers: We considered 1171 observations after trimming (26.8 % of the data was trimmed). The selected model (see Table 8) showed that the random effects are significant. Among the fixed effects, posttest score and its interaction with the pretest had the highest significance (*Estimate*= 0.0605, *SD*= 0.0237, *t*=2.56, *p*=0.012). See Supplementary Material for extensive results. Figure 6 shows the effect of the interaction between pre- and posttest score. Note that in this experiment our sample was biased towards subjects with already high scores in the pretest (this explains why the confidence interval bands in the figure become broader in those. Nonetheless, we observe a similar pattern as for the L2 German data. We observe that the RT of subjects with a low pretest, but high posttest score was low (thus they reacted fast). The RT of subjects with high pre-test scores does not seem to be affected by posttest score.

Table 8: The model with the best fit for L2 Spanish data (Experiment 2)

```
## Random effects:
##  Groups          Name        Variance Std.Dev.
##  List:Subject    (Intercept) 0.0651   0.255
##  List:Type:Trial (Intercept) 0.0244   0.156
##  Residual                    0.1572   0.396
## Number of obs: 1171, groups:  List:Subject, 100; List:Type:Trial, 32
##
## Fixed effects:
##                     Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)           8.6006     1.4741   99.0000    5.83  6.8e-08 ***
## AECI                  0.2886     0.2520  100.0000    1.15    0.255
## Post                 -0.2754     0.1461  100.0000   -1.88    0.062 .
## Pre                  -0.4004     0.2298   97.0000   -1.74    0.085 .
## TypeMISMATCH          0.6392     0.3592 1085.0000    1.78    0.075 .
## AECI:Pre             -0.0515     0.0302  101.0000   -1.70    0.092 .
## Post:Pre              0.0605     0.0237   98.0000    2.56    0.012 *
## Post:TypeMISMATCH    -0.0693     0.0379 1057.0000   -1.83    0.068 .
## Pre:TypeMISMATCH      0.0185     0.0109 1058.0000    1.69    0.091 .
```

Figure 6: the interaction between pre- and posttest scores for L2 Spanish (Experiment 2)



5. General discussion and conclusions

In this study we examined whether adult L2 learners in a foreign language context rely on grounded simulations when comprehending their L2. In particular, we investigated how L2 learners simulated the object properties of orientation and size, when reading sentences with language-specific verbs and suffixes. We instructed learners on the meaning of the verbs and suffixes and investigated if and how they simulated object orientation and size in comparison with L1 speakers. Moreover, we explored

whether individual differences in L2 learners´ proficiency, prior knowledge about the target forms, L2 use and motivation affected whether and how they made simulations.

Simulating in L1 and L2

Our first hypothesis posed that L1 and L2 German speakers would simulate object orientation whereas L1 and L2 Spanish speakers would simulate object size. We found no evidence (no match effect) for German orientation simulations in neither L1 nor L2 speakers, but we did find a match effect for Spanish size, in both L1 and L2 speakers, suggesting that both groups were making size simulations.

There are at least two accounts to explain the null effect for orientation and the match effect for size. First, both the picture and the linguistic stimuli may have differed in their salience across the two experiments. Considering the picture stimuli, Connell (2005, 2007) points out that psychophysically, properties such as shape or size, are more salient than orientation. Zwaan and Pecher (2012) also observe that orientation simulations are less robust than shape simulations and Rommers et al. (2013) found no results for orientation (but see Zwaan, 2014). Considering our linguistic stimuli, Talmy (2000) has argued that semantic components expressed in main verbs are generally backgrounded and attract limited direct attention. Talmy (2008) also poses that a referent´s divergence from certain norms tends to foreground it. Such norms and deviations from them include ordinariness versus unusualness. For example, a more unusual referent, such as a huge, heavy object (e.g. a huge, heavy book, flashlight or lipstick) tends to attract greater attention than a more ordinary referent, such as an object in prototypical size (e.g. a prototypically sized book, flashlight or lipstick). These notions suggest that German verbs may have been less salient than Spanish suffixes, which may explain the match effect for size and the null effect for orientation.

Second, we may consider task demands. Remember that participants had to answer the question ´Was this object mentioned in the preceding sentence Yes or No´. To complete the task effectively, participants only had to focus on the object nouns when reading the preceding sentences. If German participants applied such a strategy (which some participants reported) they might not have processed the orientation cue on the verb (i.e. *legen* [lay] or *stellen* [stand]). In contrast, it is likely that Spanish participants did process the size cue, since it was morphologically attached to the critical object nouns.

There is another explanation for the orientation null effect that relates to task demands. Remember that our participants gave answers by pressing the Q(Yes) or P(No) buttons on a keyboard. Pressing buttons on a keyboard involves up-and downwards motion of fingers to a key (on a vertical axis) and coordination of left and right finger pressing a left or right key (on a horizontal axis). The Theory of Event Coding (Hommel et al., 2001) predicts that when one motor response is completed (such as running a vertical orientation simulation needed to comprehend the sentence), the (vertical) feature will be bound to that outcome, making it temporarily unavailable (or less available) to other responses (such as the downward motion needed for the button press). Under such circumstances, priming between

motor responses may be eliminated or reversed (Hommel et al., 2001). When priming is eliminated, a null effect may appear in a task as ours.

Our third hypothesis posed that L1 speakers would process meaning faster than L2 speakers. Indeed, both in Experiment 1 and 2, L1 speakers showed faster RTs than L2 learners for both match and mismatch trials. This supports the idea that L2 learners may have weaker FMCs than L1 speakers (VanPatten, Williams and Rot, 2004). Slower L2 RTs might also be explained by involuntarily co-activation of L1 words during L2 sentence comprehension (Dijkstra, 2005).

Individual differences and L2 simulation

Our second hypothesis posed that differences in L2 learner factors would affect how fast L2 learners process meaning. In particular, we investigated whether differences in L2 learners´ proficiency, prior knowledge about the target forms, L2 use and motivation affected their speed of reaction.

Considering proficiency, we found an effect in Experiment 1, where more proficient learners reacted faster to stimuli than less proficient learners. No such effect was found in Experiment 2. This may be due to the larger amount of data points included in Experiment 1 as compared to Experiment 2. The smaller amount of data points in Experiment 2 may have led to less power to detect an effect in Experiment 2. As for L2 knowledge, we found that in both Experiment 1 and 2 students with lower scores in the pre- and posttest responded the slowest to critical stimuli. Arguably, students with higher pre- and posttest scores may have had prior knowledge of the target language forms before participating in the study. Potentially, this knowledge was consolidated by instruction, thereby leading to more solid FMCs, and thus faster RTs. Finally, in both Experiment 1 and 2, we found no significant effects of L2 use and motivation on performance.

Null results for L2 use and motivation indicate that such effects may be non-existent. However, they may also be explained by the nature of the current datasets. First, we observe that mainly high motivated L2 learners participated in our experiments. The lack of learners at the lower end of the scale may have taken away power to reveal significant differences. Second, we observe that the L2 Use data were skewed towards the lower end of the scale, which means that our learners did not use their L2 for many hours per day. This is not surprising given that we tested students in a foreign language context. These L2 students may have difficulties in finding L1 speakers in their environment to interact with on a daily basis. In the questionnaire (Gullberg & Indefrey, 2003) L2 German speakers reported to speak their L2 0,5 hours per day and L2 Spanish speakers 0,6 hours per day. They reported to speak the L2 with relatives, partners, friends, colleagues, classmates, roommates, but did not report whether these conversation partners were L1 speakers or not. Further, both in Germany and Spain, movies and TV-series are dubbed in German and Spanish respectively (Whitman-Linsen, 1992), which takes away opportunities for L2 students to hear their L2 outside of class. It is plausible that the lack of variance in the data we collected in a foreign language context decreased the chance to detect an effect for L2 Use.

Limitations

An important limitation in this study is that the linguistic marking made object orientation and size fully explicit and arguably relevant in completing the sentence-picture verification task. Remember the critical question in trials was ´Was this object mentioned in the preceding sentence Yes or No?´. If a sentence explicitly described a large lipstick, whereas a small lipstick is shown on the picture after, the picture invites a No response rather than a Yes response. This could induce response uncertainty leading to long response latencies. In this sense our study diverges from the simulation studies conducted by Zwaan and colleagues, where object properties (orientation, shape, color) were only implied by sentence context and irrelevant to complete the sentence-picture verification task. It is plausible that for some participants the linguistic marking induced No responses to mismatch trials. In fact, this was shown by considerable amounts of No responses to mismatch trials. However, as we only analyzed Yes responses, and removed extreme outliers from the dataset, we argue that our results are still comparable with previous simulation studies.

Another limitation is that data trimming procedures resulted in considerable loss of data. We followed trimming procedures designed by Stanfield and Zwaan (2001) and Zwaan and Pecher (2012). First, this involved the appliance of arbitrary cutoffs, which meant that original means and standard deviations were replaced by faster means with lower variance. Second, a focused trimming procedure was applied, where means and standard deviations were calculated for each participant. RTs further away than two standard deviations from individuals´ means were trimmed. Finally, we analyzed median RTs. These substantial procedures decrease how reflective results are as compared with the natural data. Yet, comparability with the many previous simulation studies was of crucial importance to us and outweighed the limitations the trimming procedures brought along. Thus, after consideration, we chose to keep trimming procedures comparable with previous studies.

Conclusions

The key finding in this study is that both L1 and L2 learners have been found to make mental simulations in relation to the property of objects´ size. Thus, size matters. When given information about object size in language, both L1 and adult L2 speakers in a foreign language context rely on mental simulations rather than amodal symbols when comprehending language. We consider this finding as a first step towards the use of monolingual simulation theory (Barsalou, 1999) to benefit a model that can explain L2 mental simulation in L2 comprehension. Such a model would need to take into account that L2 simulation speed may be affected by at least the following three aspects. First, simulation speed is slower in an L2 as compared with L1, possibly due to weaker L2 FMCs or co-activation of L1 forms. Second, speed of L2 simulation is affected by L2 learner factors. In this paper we found partial support for the notion that proficiency would need to be considered. In addition, knowledge of particular FMCs (through L2 instruction) may affect how fast L2 learners process meaning. Third, following our interpretation of results, a comprehensive model of L2 simulation may need to take into account how physical task-demands affect mental activity.

Future studies need to consolidate the knowledge generated in this study and investigate if and how other object properties are being simulated in an L2 and what factors need to be accounted for. In addition, it is critical to investigate whether L2 learners simulate in extended discourse (for L1 research, see Ditman et al., 2010). Such expansions on monolingual theory on language comprehension may feed into existing models of second language comprehension (Dijkstra, 2005; Thomas & Van Heuven, 2005) or lead to the formulation of new hypotheses on the mechanisms and processes at work in second language comprehension.

Author contributions:
Author 1, 2 and 3 designed this study. Testing and data collection were performed by Author 1. Author 1 and 4 determined the type of analyses. Author 4 performed data analysis. Author 1 and 4 determined how to present results. Author 1 drafted the manuscript and Author 2, 3 and 4 provided critical revisions. All authors approved the final version of the manuscript for submission.

Acknowledgements:
This work has been supported by the EU 7th Framework Programme Marie Curie Initial Training Networks grant Nr. 316748 under the project *Language and Perception*. We thank our collaborators, hosts and local assistants who enabled data collection at the various test locations.

End notes:
1. In the pilot study 12 participants were asked to consider 112 object combinations. They were asked how they would place the first object in each combination onto the second object. They responded by ticking a box on a five point scale that ranged from 1, "lying" to 3, "either way" to 5, "standing". For purposes of the current study we selected four object combinations that received an average rating of 3, "either way".

Supplementary Material:
Available through https://github.com/belzebuu/LanguageStudy.

References

Baddeley, A. (1990). Human memory: Theory and practice. Needham Heights, MA: Allyn & Bacon.

Bardovi-Harlig, K. (2000). Tense and aspect in second language acquisition: Form, meaning and use. Language Learning, 50 (Supplement 1).

Barsalou, L. (1999a). Language comprehension: Archival memory or preparation for situated action? *Discourse Processes*, 28, 61-80.

Barsalou, L. (1999b). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-660.

Barsalou, L. (2008). Grounded Cognition. *Annual Review of Psychology* 59, 617-645.

Bartoń, K. (2016). MuMIn: Multi-Model Inference. R package version 1.15.6.

Bates, D., Machler, M., Bolker B., Walker, S. (2015). Fitting linear mixed-effects modeling using lm4. *Journal of Statistical Software* 67(1), 1-48.

Benati, G., VanPatten, B. (2010). *Key terms in second language acquisition*. London, GBR: Continuum International Publishing.

Bergen, B., Wheeler, K. (2010). Grammatical aspect and mental simulation. *Brain & Language*, 112, 150-158.

Butt, J., Benjamin, C. (2005). *A new reference grammar of Modern Spanish*. Fourth edition. London: Arnold.

Burgess, C., Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, 12, 177-210.

Collentine, J., Freed, B. (2004). Learning context and its effects on second language acquisition. Introduction. SSLA, 26, 153-171.

Colunga, E. & Smith, L.B. (2005) From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, 112, 347-382.

Connell, L. (2005). Colour and stability in embodied representations. In B. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*, pp. 482-487, Mahwah, NJ: Lawrence Erlbaum.

Connell, L. (2007). Representing colour in language comprehension. *Cognition* 102, pp. 476-485.

De Groot, A. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 1001-1018.

De Knop, S., Perrez, J. (2014). Conceptual metaphors as a tool for the efficient teaching of Dutch and German posture verbs. *Review of Cognitive Linguistics*, 12(1), 1-29.

Ditman, T., Brunyé, T., Mahoney, C., Taylor, H. (2010). Simulating an enactment effect: Pronouns

guide action simulation during narrative comprehension. *Cognition,* 115, 172-178.

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. In Kroll, J., De Groot, A. (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 179-201, Oxford: Oxford University Press.

Ellis, N. (Ed.) (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, 24(2), 143-188.

Ellis, N. (2004). The process of second language acquisition. In VanPatten, B., Williams, J., Rott, S. (Eds.), *Form-meaning connections in second language acquisition*, pp. 49-76, Lawrence Erlbaum Associates Publishers, Londen, Mahwah: New Jersey.

Ellis, N., (2005). Introduction to Part I: Acquisition. In Kroll, J., De Groot, A. (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 3-8, Oxford: Oxford University Press.

Fagan, S. (1991). The Semantics of the Positional Predicates liegen/legen, sitzen/setzen, and stehen/stellen. *Die Unterrichtspraxis / Teaching German*, Wiley: American Association of Teachers of German, 24, 136-145.

Feldman, J. (2006). *From molecule to metaphor: A neural theory of language*. Cambridge: MIT Press.

Glenberg, A. (1997). What memory is for. *Behavioral & Brain Sciences*, 20, 1-55.

Glenberg, A., Robertson, D. (1999). Indexical understanding of instructions. *Discourse Processes*, 28, 1-26.

Glenberg, A., Robertson, D. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory & Language*, 43, 379-401.

Gullberg, M., Indefrey, P. (2003). Language background questionnaire: Developed in the dynamics of multilingual processing project. Nijmegen: Max Planck Institute for Psycholinguistics.

Gooch, A. (1967). *Diminutive, augmentative and pejorative suffixes in modern Spanish. A guide to their use and meaning*. Oxford: Pergamom Press.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*, 849-878.

Hualde, J., Olarrea, A., Escobar, A., Travis, C. (2010) *Introducción a la lingüística hispánica*. 2nd. Edition. Cambridge: Cambridge University Press.

Hymes, D. (1972). Reinventing anthropology. New York: Random House.

Indefrey, P. (2006). A meta-analysis of hemodynamic studies on first and second language processing: Which suggested differences can we trust and what do they mean? *Language Learning*, 56, 279-304.

Izura, C., Ellis, A.W. (2002). Age of acquisition effects in word recognition and production in first and second languages. *Journal of Memory and Language,* 50, 165-181.

Kaschak, M., Glenberg, A. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. Journal of Memory and Language, 43, 508-529.

Korecky-Kröll, K., Dressler, W. Diminutives and hypocoristics in Austrian German (AG). In Savickiene, I. & Dressler, W. (Eds.), *Acquisition of diminutives: A cross-linguistic perspective*, pp. 207-230, Amsterdam: John Benjamins Publishing Company.

Kuznetsova, A., Brockhoff, P., Christensen, R. (2016). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-30.

Lemmens, M. (2006). Caused posture: Experiental patterns emerging from corpus research. Trends in Linguistics, Studies and Monographs: Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis. Walter de Gruyter.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Marcos Miguel, N. (2010). Grapho-morphological awareness in Spanish L2 reading. Unpublished master thesis. University of Pittsburgh.

Norris, J., Ortega, L. (2001). Effectiveness of L2 instruction: a research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417-528.

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

Pulvermuller, F. (2012) Meaning and the brain: The neurosemantics of referential, interactive, and combinatorial knowledge. *Journal of Neurolinguistics*, 25(5), 423-459.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rommers, J., Meyer, A., Huettig, F. (2013) Object shape and orientation do not routinely influence performance during language processing. *Psychological science,* published online, 1-8.

Sato, M., Schafer, A., Bergen, B. (2013). One word at a time: Mental representations of object shape change incrementally during sentence processing. *Language and Cognition,* 5(4), 345-373.

Shirai, Y. & Andersen, R. (1995). The acquisition of tense-aspect morphology. Language, 71, 743-762.

Stanfield, R.A., Zwaan, R. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science,* 12, 153-156.

Talmy, L. (2000). Chapter 1: Lexicalization patterns. In *Toward a cognitive semantics, volume 2: typology and process in concept structuring*, pp. 21-146, Cambridge, MA: MIT Press 2000.

Talmy, L. (2013). Concept structuring systems in language. In M. Tomasello (Ed.), The new psychology of language: Cognitive and functional approaches to language structure (pp. 15–46). Mahwah, NJ: Lawrence Erlbaum.

Thomas, M., Van Heuven, W. (2005). Computational models of bilingual comprehension. In: Kroll, J., De Groot, A. (Eds.): *Handbook of bilingualism: Psycholinguistic approaches,* pp. 202-225, Oxford: Oxford University Press.

Tokowicz, N., Perfetti, C. (2005). Introduction to Part II: Comprehension. In Kroll, J., De Groot, A. (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 173-178, Oxford: Oxford University Press.

Tomzcak, E., Ewert, A. (2015). Real and Fictive Motion Processing in Polish L2 Users of English and Monolinguals: Evidence for Different Conceptual Representations. *The Modern Language Journal*, 99, 49-65.

VanPatten, B., Williams, J., Rott, S. (2004). Form-meaning connections in second language acquisition. In VanPatten, B., Williams, J., Rott, S. (Eds.), *Form-meaning connections in second language acquisition*, pp. 1-26. Lawrence Erlbaum Associates Publishers, Londen, Mahwah: New Jersey.

Veivo, O., Jarvikivi, J. (2013). Proficiency modulates early orthographic and phonological processing in L2 spoken word recognition. *Bilingualism: Language & Cognition*, 16(4), 864-883.

Vukovic, N., Williams, J. (2014). Automatic perceptual simulation of first language meanings during second language sentence processing in bilinguals. *Acta Psychologica* 145, 98-103.

Vukovic, N., Williams, J. (2015). Individual differences in spatial cognition influence mental simulation of language. *Cognition*, 142, 110-122.

Whitman-Linsen, C. (1992) *Through the Dubbing Glass: the Synchronization of American Motion Pictures into German, French and Spanish.* Frankfurt am Main: Peter Lang.

Zwaan, R., Stanfield, R., Yaxley, R. (2002). Language comprehenders mentally represent the shape of objects. *Psychological Science*, 13, 168-171.

Zwaan, R., Pecher, D. (2012). Revisiting mental simulation in language comprehension: six replication attempts. *PloS One*, 7(12), published online, 1-10.

Zwaan, R. (2014). Replications should be performed with power and precision: A response to Rommers, Meyer, and Huettig (2013). *Psychological Science*, 25, 305-307.

Appendices

Appendix 1: instruction on the German verbs *legen/stellen* (in German)

Die Verben *stellen, legen, liegen* und *stehen*

Wussten Sie schon?
Auf Deutsch gibt es Verben, die die Position von Objekten oder Personen beschreiben. Diese Verben zeigen an ob das Objekt oder die Person sich in einer vertikalen oder horizontalen Position befindet. Vier Verben, die oft benutzt werden, sind:

1. stellen poner en una posición vertical
2. legen poner en una posición horizontal
3. liegen estár en una posición horizontal
4. stehen estár en una posición vertical

Falls man die Position oder Bewegung noch genauer beschreiben möchte, kann man im Deutschen noch *gerade-* oder *hin-* hinzufügen (zum Beispiel: *geradestellen, geradelegen; hinstellen, hinlegen*).

Konjugationen

stellen	**stehen**	**legen**	**liegen**	
Ich	stelle	stehe	lege	liege
Du	stellst	stehst	legst	liegst
Er/sie	stellt	steht	legt	liegt
Wir	stellen	stehen	legen	liegen
Ihr	stellt	steht	legt	liegt
Sie	stellen	stehen	legen	liegen

Beispiel

Ulrich: "Uf, dieses Buch ist sehr groß und schwer!"

Kathi: "Du Armer! Komm, dann **legen** wir das auf den Tisch!"

Appendix 2: vocabulary list with 30 L2 German words and their Spanish translation. The original document had standard typeface, but for information purposes, here, we indicate the fifteen target words in bold.

Vokabeln

in die Experimente werden Sie eine Serie deutsche Worte sehen. Es ist wichtig dass Sie verstehen was diese Worten bedeuten. Hierunter finden Sie eine Wortliste mit spanischer Übersetzung. Bitte, benutzen Sie 5 Minuten um diese Worten zu lehren. Wenn Sie bereit sind, können Sie anfangen mit dem ersten Experiment.

1.	die Treppe	la escalera
2.	der Ball	la pelota
3.	das Lineal	la regla
4.	die Tomate	el tomate
5.	**das Schneidebrett**	**la tabla de cortar**
6.	**die Taschenlampe**	**la linterna**
7.	**das Papier**	**el papel**
8.	die Trompete	la trompeta
9.	die CD	el CD
10.	**die Tube**	**el tubo**
11.	der Stuhl	la silla
12.	**der Klebestift**	**la barra de pegamento**
13.	die Säge	la sierra
14.	**das Deodorant**	**el desodorante**
15.	die Waschmaschine	la lavadora
16.	**die Batterie**	**la pila**
17.	die Kerze	la vela
18.	der Stern	la estrella
19.	das Fernglas	el par de binoculares
20.	**das Regal**	**el estante**
21.	**das Handtuch**	**la toalla**
22.	**der Lippenstift**	**la barra de labios**
23.	die Pfanne	la sartén
24.	**der Tisch**	**la mesa**
25.	der Stift	el lápiz
26.	**die Glocke**	**la campana**
27.	**die Bandspule**	**el carrete del hilo**
28.	**das Brett**	**el plato**
29.	der Schrank	el armario
30.	**die Matte**	**la estera**

Complete References:

Abu-Rabia, S. (2003). The influence of working memory on reading and creative writing processes in a second language. *Educational Psychology*, 23, 209-219.

Abutalebi, J., Cappa, S., Perani, D., (2005). What can functional neuroimaging tell us about the bilingual brain? In: Kroll, J., de Groot, A. (Eds.) *Handbook of Bilingualism. Psycholinguistic Approaches.* (pp.497-516). Oxford University Press.

Ainciburu, M. (2011). *Vía rapída A1-B1*. Madrid: Difusion Centro de Publicacion y Publicaciones de Idiomas, S.L.

Akaike, H. (1987). Factor analysis and the AIC. *Psychometrika*, 52, 317-332.

Alarcos Llorach, E. (1994). *Gramática de la lengua española*. Real Académica Española, Colección Nebrija y Bello. España: Artes Gráficas Penalara.

Alferink, I., Gullberg, M. (2013). French–Dutch bilinguals do not maintain obligatory semantic distinctions: Evidence from placement verbs. *Bilingualism: Language and Cognition*, 17(1), 22-37.

Álvarez Martínez, A. (1989). *El pronombre. Personales, Artículo, Demonstrativos, Posesivos*. España: Grafur, S.A.

Ameka, F. & Levinson, S. (Eds). (2007). The typology of semantics of locative predication: Posturals, positionals and other beasts. Special Issue of *Linguistics* 45. Berlin: Mouton de Gruyter.

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., Riches, N. & Thierry, G. (2015). Two languages, two minds: flexible cognitive processing driven by language of operation. *Psychological Science*. 26, 4, 518-526.

Athanasopoulos, P., Bylund, E. (2013). The 'thinking' in thinking-for-speaking: where is it? In : *Language, Interaction et Acquisition / Language, Interaction and Acquisition*, 4(1), 91-100.

Babbie, E. (2015). *The practice of social research*. Stamford, CT: Cengage Learning.

Baker, C. (1993). *Foundations of Bilingual Education and Bilingualism*. Clevedon: Multilingual Matters.

Baker, C., Prys-Jones, S. (1998). *Encyclopedia of Bilingualism and Bilingual Education*. Clevedon: Multilingual Matters.

Barsalou, L. (1999a). Language comprehension: Archival memory or preparation for situated action? *Discourse Processes, 28*, 61-80.

Barsalou, L. (1999b). Perceptual symbol systems. *Behavioral and Brain Sciences,* 22, 577-660.

Barsalou, L. (2008). Grounded Cognition. *Annual Review of Psychology*, 59, 617-645.

Bassetti, B., Cook, V. (2011) Language and cognition: The second language user. In V. Cook and B. Bassetti (Eds.) *Language and Bilingual Cognition*. (pp. 143-190). Oxford, UK: Psychology Press.

Bergen, B., Wheeler, K. (2010). Grammatical aspect and mental simulation. *Brain & Language*, 112, 150-158.

Berman, R., Slobin, D. (Eds.) (1994). *Relating events in narrative: A crosslinguistic developmental study* (Vol. 1). Hillsdale: NJ: Lawrence Erlbaum Associates.

Beyn, E., Knyazeva, G. (1962). The problem of prosopagnosia. *The Journal of Neurology, Neurosurgery, and Psychiatry*, 25, 154-158.

Billman, D. Swilley, A., Krych, M. (2000). Path and manner priming: verb production and event recognition. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society.* (pp. 615-620). Mahwah, NJ: Lawrence Erlbaum.

Boroditsky, L., Ham, W., Ramscar, M., (2002). What is universal in event perception? Comparing English and Indonesian speakers. *Proceedings of the 24th annual meeting of the Cognitive Science Society.* (pp. 136-141). Mahwah, NJ: Lawrence Erlbaum.

Boroditsky, L., Schmidt, L., & Phillips, W. (2003). Sex, Syntax, and Semantics. In Gentner & Goldin-Meadow (Eds.,) *Language in Mind: Advances in the study of Language and Cognition.* Cambridge, MA: MIT Press.

Borregine, K., Kaschak, M. (2006). The action compatibility effect: it´s all in the timing. *Cognitive Science*, 30, 1097-1112.

Bosse, S., Papafragou, A. (2010). Spatial position in Language and Visual Memory: A Cross-Linguistic Comparison. *Proceedings from the 32nd Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Bosque, I. (2010). *Nueva gramática de la lengua española. Manual*. Real academia española. Asociación de academias de la lengua española. España: Espasa Libros S.L.U.

Bosque Muñoz, I., Demonte Barreto, V. (1999). *Gramática descriptiva de la lengua española. Sintaxis básica de las clases de palabras.* España: Rotapapel, S.L.

Boutonnet, B., Athanasopoulos, P., Thierry, G. (2012). Unconscious effects of grammatical gender during object categorization. *Brain Research*, 1479, 72-79.

Bowerman, M. (1980). The structure and origin of semantic categories in the language learning child. In M. Foster, & S. Brandes (Eds.), *Symbol as sense* (pp. 277-299). New York: Academic Press.

Brown, R., Lenneberg, E. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology*, 49, 454-462.

Brysbaert M, Duyck W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after 15 years of service? *Bilingualism: Language and Cognition,* 13(3), 359–371.

Buchler, N., Reder, L. (2007). Modeling age-related memory deficits: a two-parameter solution. *Psychology and Aging*, 22, 104-121.

Burgess, C., Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, 12, 177-210.

Butt, J., Benjamin, C. (1998). *A new reference grammar of modern Spanish*. New York: Oxford University Press.

Butt, J., Benjamin, C. (2005). *A new reference grammar of modern Spanish*. London: Arnold.

Cadierno, T., Ibarretxe-Antunano, I., Hijazo-Gascón, A. (2016). Semantic categorization of placement verbs in L1 and L2 Danish and Spanish. *Language Learning*, 66(1), 191-223.

Cadierno, T. (1995). Formal instruction from a processing perspective: An investigation into the Spanish past tense. *The Modern Language Journal*, 79, 2, 179-193.

Carroll, D. (1994). *Psychology of language.* Pacific Groove, CA: Brooks and Cole.

Carroll, J. (1956). Introduction. In J. Carroll (ed.) *Selected writings of Benjamin Lee Whorf*. (pp. 1-34) Cambridge: MIT Press.

Cartagena, N., Gauger, H. (1989). *Spanisch-Deutsch 1. Duden vergleichende Grammatiken.* Mannheim: Druckerei Krembel, Spreyer.

Carranza, R., Seguí, V. (2005). La productividad de la sufijación apreciativa en el habla de Córdoba (Argentina). In M. Carballo (Ed.): *Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua, deseo y realidad: Actas del XV Congrese Internacional de ASELA*, Seville, September 22-25, 199-206.

Casad, E., Langacker, R. (1985). "Inside" and "Outside" in core grammar. *International Journal of American Linguistics,* 51(1), 247-281.

Cardini, F. (2010). Evidence against Whorfian effects in motion conceptualization. *Journal of Pragmatics,* 42, 1442–1459.

Cenoz, J. (2001). The effect of linguistic distance, L2 status and age on cross linguistic influence in third language acquisition. In: J. Cenoz, B. Hufeisen & U. Jessner (Eds), *Cross linguistic influence in third language acquisition: Psycholinguistic perspectives* (pp. 8-20). Clevedon, UK: Multilingual Matters.

Chomsky, N. (1980). *Rules and Representations*. Colombia University Press.

Connell, L. (2005). Colour and stability in embodied representations. In B. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*. (pp. 482-487). Mahwah, NJ: Lawrence Erlbaum.

Connell. L. (2007). Representing colour in language comprehension. *Cognition,* 102, 476-485.

Cook, V., Bassetti, B. (Eds.) (2011). *Language and Bilingual Cognition*. New York: Psychology Press.

Coventry, K. R., Valdés, B., & Guijarro-Fuentes, P. 2010. Thinking for speaking and immediate memory for spatial relations. In Z.-H. Han & T. Cadierno (Eds.), *Linguistic Relativity in L2 Acquisition: Evidence of L1 Thinking for Speaking. (*pp. 84-101). Clevedon: Multilingual Matters.

Crookes, G., Schmidt, R. (1991). Motivation: Reopening the research agenda. *Language Learning,* 41(4), 469-512.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science,* 25, 7-29.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

De Bruyne, J. (1993). *Spanische Grammatik*. Deutschland: Franz Spiegel Buch.

Dijkstra, T. (2005). Bilingual Visual Word Recognition and Lexical Access. In: Kroll, J. & De Groot, J.

(Eds.), *Handbook of Bilingualism. Psycholinguistic Approaches*. (pp. 173-178). Oxford University Press.

Dijkstra, A., Van Heuven, W. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition,* 5, 175-197.

Ditman, T., Brunyé, T., Mahoney, C., Taylor, H. (2010). Simulating an enactment effect: Pronouns guide action simulation during narrative comprehension. *Cognition*, 115, 172-178.

Dornyei, Z., Skehan, P. (2003). Individual differences in second language learning. In C. Doughty & M. Long (Eds.) *The handbook of second language acquisition* (pp. 589-630). Malden, Oxford, Carlton: Blackwell Publishing Ltd.

Drosdowski, G. (1995). *Duden. Die Grammatik. Unentbehrlich für richtiges Deutsch*. 4. Neuauflage. Berchtesgaden: Graphische Betriebe Langenscheidt.

Durgin, F., Baird, J., Greenburg, M., Russel, R., Shaughnessy, K., Waymouth, S. (2009). Who is being deceived? The experimental demands of wearing a backpack. *Pscyhnomic Bulletin & Review*, 16, 964-969.

Dufour, R., Kroll, J.F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition,* 23(2), 166-180.

Ellis, N. (2004). The process of second language acquisition. In: VanPatten, B., Williams, J., Rott, S. (Eds.), *Form-meaning connections in second language acquisition* (pp. 49-76), Lawrence Erlbaum Associates Publishers, London, Mahwah: New Jersey.

Ellis, N. (2006). Selective attention and transfer phenomena in SLA: Contingency, cue competition, salience, interference, overshadowing, blocking and perceptual learning. *Applied Linguistics*, 27(2), 1-31.

Ellis, N., (2008). The dynamics of language use, language change, and first and second language acquisition. *Modern Language Journal*, 41(3), 232-249.

Ellis, N. (2008). Usage-based and form-focused language acquisition: The associative learning of constructions, learned attention, and the limited L2 endstate. In P. Robinson & N. Ellis (Eds.): *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 372-405). New York, Abingdon: Routledge.

Ellis, N., Cadierno, T. (2009). Constructing a second language: Introduction to the Special section. Annual Review of Cognitive Linguistics, 7, 111-139.

Ellis, R. (1994). *The study of second language acquisition.* Oxford, UK: Oxford University Press.

Ellis, R. (2001). Investigating form-focused instruction. *Language Learning*, 51(1), 1- 46.

Engelen, J., Bouwmeester, S., de Bruin, A., Zwaan, R. (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology*, 110, 659-675.

Ervin-Tripp, S. (1954, reprinted in 1973). Identification and bilingualism. In A. Dil (ed.). *Language acquisition and communicative choice. Essays by Susan M. Ervin-Tripp*. (pp. 45-61). Stanford, CA: Stanford University Press.

Ervin-Tripp, S. (1967, reprinted in 1973). Semantic shift in bilinguals. In A. Dil (ed.). *Language acquisition and communicative choice. Essays by Susan M. Ervin-Tripp*. (pp. 33-44). Stanford, CA: Stanford University Press.

Fagan, S. (1991). The Semantics of the Positional Predicates liegen/legen, sitzen/setzen, and stehen/stellen. *Die Unterrichtspraxis / Teaching German*, 24, 136-145.

Feist, M., Gentner, D. (2007). Spatial language influences memory for spatial scenes. *Memory & Cognition*, 35(2), 283-296.

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. 4th Edition. CA: Sage Publications

Filipović, L. (2011). Thinking and remembering in one or two languages: bilingual vs. monolingual lexicalization and memory for motion events. *International journal of bilingualism*, 15(4), 466-485.

Finkbeiner, M., Nicol, J., Greth, D., Nakamura, K. (2002). The role of language in memory for actions. *Journal of Psycholinguistic Research*, 31(5), 447- 457.

Firestone, C., Scholl, B. (2014). "Top-down" effects where none should be found: The El Greco fallacy in perception research. *Pscyhological Science*, 25(1), 38-46.

Firestone, C., Scholl, B. (in press). Cognition does not affect perception: Evaluating the evidence for ' top-down' effects. *Behavioral & Brain Sciences.*

Fischer, M., Zwaan, R. (2008) Embodied language – a review of the role of the motor system in language comprehension. *Q. J. Exp. Psychol*. 61, 825–850

Fodor, J. (2000). *The mind doesn´t work that way: the scope and limits of computational psychology*. Cambridge: MIT Press.

Francis, W., Gutierrez, M. (2012). Bilingual recognition memory: Stronger performance but weaker levels-of-processing effects in the less fluent language. *Memory and Cognition*, 40, 496-503.

Francis, W., Strobach, N. (2013). The bilingual L2 advantage in recognition memory. *Psychonomic Bulletin and Review*, 20, 1296-1303.

Gardner, H. (1985). *The minds new science. A history of the cognitive revolution.* New York: Basic Books, Inc. Publisher.

Gardner, R. (1982). Language attitudes and language learning. In E. Bouchard Ryan & H. Giles, *Attitudes towards language variation* (pp. 132-147). Edward Arnold.

Gardner, R., & Lambert, W. (1972). *Attitudes and motivation : Second language learning*. Newbury House.

Gardner, R. (1992). Second Language Learning in Adults: Correlates of Proficiency. *Applied Language Learning*, 2, 1-28.

Gennari, S., Sloman, S., Malt, B., Fitch, W. (2002). Motion events in language and cognition. *Cognition,* 83, 49-79.

Gentner, D., Goldin-Meadow, S. (2003). Whither Whorf. In Gentner, D. & Goldin-Meadow (Eds.): *Language in Mind. Advances in the Study of Language and Thought.* Cambridge, London, Massachusetss: MIT Press.

Glatz, D., Von Stutterheim, C. (2003). Introduction. In Glatz, D. & Von Stutterheim, C. (Eds.): Special issue: Perspectives in language production. Linguistics, 41(5), Berlin, New York: Mouton de Gruyter.

Gleitman, L., Papafragou, A. *(*2005*).* Language and thought. In K. Holyoak & R. Morrison (Eds.), *Cambridge Handbook of Thinking and Reasoning* (pp. 633-661), Cambridge University Press.

Glenberg, A. (1997). What memory is for. *Behavioral & Brain Sciences*, 20, 1-55.

Glenberg, A., Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558-565.

Glenberg, A., Robertson, D. (1999). Indexical understanding of instructions. *Discourse Processes*, 28, 1-26.

Glenberg, A., Robertson, D. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory & Language*, 43, 379-401.

Goldschneider, J., DeKeyser, R., (2001). Explaining the „natural order of L2 morpheme acquistion“ in English: A meta-analysis of multiple determinants. *Language Learning*, 51, 1-50.

Gollan, T., Montoya, R., Cera, C., Sandoval, T. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, 58, 787-814.

Gooch, A. (1967). *Diminutive, augmentative and pejorative suffixes in modern Spanish. A guide to their use and meaning*. Oxford: Pergamom Press.

Grosjean, F. (*1982*). *Life with Two Languages: An Introduction to Bilingualism.* Cambridge, Mass: Harvard University Press.

Hae in Park, H., Ziegler, N. (2014). Cognitive shift in the bilingual mind: Spatial concepts in Korean–English Bilinguals. *Bilingualism: Language and Cognition*, 17(2), 410–430.

Han, Z., Cadierno, T. (2010). *Linguistic relativity in SLA: Thinking for speaking.* Bristol, Buffalo, Toronto: Multilingual Matters.

Harnad, S. (*1990*). *The Symbol Grounding Problem*. Physica, 42, 335-346.

Hellinger, M., Bußmann, H. (2001, 2002, 2003). Gender Across Languages: The Linguistic Representation of Women and Men, Vol. 1, 2, 3. Amsterdam: Benjamins.

Hoijer, H. (1954). The Sapir-Whorf hypothesis. In Hoijer, H. (ed.) *Language in culture: Conference on the interrelations of language and other aspects of culture* (pp. 92-105). University of Chicago Press.

Hoeben Mannaert, L. (2016). Is it possible to update mental simulations? Presentation at Language and Perception Conference, NTNU, Trondheim, June 13-16.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24, 849-878.

Hualde, J., Olarrea, A., Escobar, A., & Travis, C. (2010) *Introducción a la lingüística hispánica.* Second Edition. Cambridge: Cambridge University Press.

Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.

Jacobovits, L., Lambert, W. (1961). Semantic satiation among bilinguals. *Journal of Experimental Psychology*, 2, 576-582.

Jarvis, Scott (Ed.) (2011). Bilingualism: Language and Cognition, 14 (1). (Special issue on crosslinguistic influence in bilinguals' concepts and conceptualizations, guest-edited by Scott Jarvis).

Jarvis, S., Pavlenko, A. (2007). *Crosslinguistic influence in language and cognition*. New York: Routledge.

Jarvis, S. (1998). Conceptual transfer in the interlingual lexicon. Bloomington: IN: Indiana University Linguistics Club Publications.

Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 Influence in the interlanguage lexicon. *Language Learning*, 50, 245-309.

Kaschak, M., Madden, C., Therriault, D., Yaxley, H., Aveyard, M., Blanchard, A., Zwaan, R. (2005). Perception of motion affects language processing. *Cognition*, 94, B79-B89.

Kecskes, I., Papp, T. (2000). *Foreign language and mother tongue.* Mahwah, NJ: Erlbaum.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge, UK: Cambridge University Press.

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: APA Books.

Kopecka, A., Narasimhan, B. (Eds). (2012). *Events of Putting and Taking, a crosslinguistic perspective*. Amsterdam: Benjamins.

Korecky-Kröll, K., Dressler, W. (2007). Diminutives and hypocoristics in Austrian German (AG). In: Savickiene, I. & Dressler, W. (Eds.) *Acquisition of diminutives: A cross-linguistic perspective*. Amsterdam: John Benjamins Publishing Company.

Kroll, J., Van Hell, J., Tokowicz, N., Green, D. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language & Cognition*, 13(3), 373/381.

Kroll, J., De Groot, A. (Eds.) (2005). *Handbook of Bilingualism. Psycholinguistic Approaches.* Oxford University Press.

Kroll, J., Tokowicz, N. (2005). Models of bilingual representation and processing: looking back and to the future. In Kroll, J., De Groot, A. (Eds.): *Handbook of Bilingualism. Psycholinguistic Approaches.* Oxford University Press, pp. 531-554.

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

Kruschke, J. K. (2010). *Doing Bayesian data analysis: A tutorial with R and BUGS*. New York, NY: Academic Press.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers.* Ann Arbor: University of Michigan Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Lambert, W., Havelka, J., Crosby, C. (1958). The influence of language acquisition contexts on bilingualism. *Journal of Abnormal and Social Psychology*, 56, 239-244.

Lambert. W., Rawlings, C. (1969). Bilingual processing of mixed-language associative networks. *Journal of Verbal Learning and Verbal Behavior*, 8, 604-609.

Langacker, R. (1976). Semantic representations and the linguistic relativity hypothesis. *Foundations of Language*, 14(3), 307-357.

Larsen-Freeman, D. (1976). An explanation for the morpheme acquisition order of second language learners. *Language Learning*, 26, 125-134.

Lázaro Mora, F. (1999). *La derivación apreciativa.* In: Ignacio Bosque & Violeta Demonte (Eds.), Gramática descriptiva de la lengua española. Madrid: Espasa Calpe.

Lemmens, M. (2006). Caused posture: Experiental patterns emerging from corpus research. In: Stefanowitsch, A. & S. Gries (Eds.) *Corpora in Cognitive Linguistics. Vol II: The Syntax-Lexis Interface*. Berlin: Mouton De Gruyter.

Levelt, W. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levine, T., Hullett, C. (2002). Eta squared, Partial Eta Squared, and Misreporting Effect Sizes in Communication Research. *Human communication research*, 28(4), 612-625.

Levinson, S. C. (1996). Introduction to part II. In J. J. Gumperz, & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 133-144). Cambridge: Cambridge University Press.

Levinson, S. 2012. Preface. In A. Kopecka, & B. Narasimhan (Eds.), *Events of putting and taking: A crosslinguistic perspective*. Amsterdam: Benjamins.

Li, P, Gleitman, L. (2002).Turning the tables: language and spatial reasoning. *Cognition*, 83, 265–294.

Loftus, E., & Palmer, J. (1974). Reconstruction of auto-mobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 13, 585-589.

Lohde, M. (2006). *Wortbildung des modernen Deutschen: ein Lehr- und Übungsbuch.* Tübingen: Gunter Narr.

Lloret Ivorra, E., Ribas, R., Wiener, B., Pérez Cañizares, P. (2010). *Con Gusto A1-A2.* Stuttgart: Ernst Klett Sprachen.

Long, M, Robinson, P. (1998). Focus on form. Theory, research and practice. In C. Doughty and J. Williams (Eds.) *Focus on form in classroom second language acquisition* (pp. 144-138). Cambridge: Cambridge University Press.

Longford, N. (1993). *Random coefficient models.* New York: Oxford University Press.

Luke, D., (2004). *Multilevel modeling.* SAGE Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage Publications.

Mackey, A., Gass, S. (2005). *Second language research: methodology and design.* Mahwah, New Jersey, London: Lawrence Erlbaum Publishers.

Macnamara, J. (1970). Bilingualism and thought. In: J. Atlis (ed.) Georgetown University 21st Annual Round Table (pp. 25-40). Washington, DC: Georgetown University Press.

Macnamara, J. (1991) Linguistic relativity revisited. In: R. Cooper and B. Spolsky (Eds.) *The influence of Language on Culture and Thought: Essays in Honor of Joshua A. Fishmans 65th Birthday* (pp. 45-60). Berlin: Mouton de Gruyter.

Martín Peris, E., Sans Baulenas, N. (2013). *Gente 1 hoy (A1-A2).* Madrid: Difusion Centro de Publicacion y Publicaciones de Idiomas, S.L.

Matlin, M. (2005). *Cognition.* US: John Wiley & Sons, Inc.

Meyer, D., Schvaneveldt, R. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*. 90, 227–234.

Miller, G., Johnson-Laird, P. (1976). *Language and Perception.* Harvard University Press.

Narasimhan, B., Kopecka, A., Bowerman, M., Gullberg, M., Majid, A. (2012). Putting and taking events: a crosslinguistic perspective. In A. Kopecka, & B. Narasimhan (Eds.), *Events of putting and taking: A crosslinguistic perspective*. Amsterdam: Benjamins.

Nieuwland, M., Van Berkum, J. (2006) When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18, 1098–1111.

Norris, J., Ortega, L. (2001). Effectiveness of L2 instruction: a research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417-528.

Nosofsky, R. (1986). Attention, Similarity and the Identification-Categorization Relationship. *Journal of Experimental Psychology: General*, 115(1), 39-57.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge, UK: Cambridge University Press.

Papafragou, A., Massey, C., Gleitman, L. (2002). Shake, rattle, 'n' roll: the representation of motion in language and cognition. *Cognition*, 84, 189–219.

Papesh, M. (2015). Just out of reach: On the reliability of the action-sentence compatibility effect. *Journal of Experimental Psychology: General*, 144(6), c116-c141.

Pavlenko, A. (2011). *Thinking and speaking in two languages*. Clevedon, England: Multilingual Matters.

Pavlenko, A. (2014). *The bilingual mind. And what it tells us about language and thought*. Cambridge University Press.

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

Pederson, E. *(*1995). Language as context, language as means: Spatial cognition and habitual language use. *Cognitive Linguistics*, 6(1), 33-62.

Perfetti, C. (1999). Comprehending written language: A blueprint of the reader. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 167-208). Oxford University Press.

Philipps, W., Boroditsky, L. (2003). Can quirks of grammar affect the way you think? Grammatical gender and object concepts. In R. Alterman & D. Kirsh (Eds.), Proceedings of the 20-fifth annual meeting of the Cognitive Science Society. Boston: Cognitive Science Society.

Pinker, *S. (*1994*). The Language Instinct.* New York, NY: Harper Perennial Modern Classics.

RAE. Real Academia Española. (2009). *Nueva gramática de la lengua Española. Morfología. Sintaxis 1.* Madrid: Espana.

Raudenbusch, S., Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd. Ed.). Thousand Oaks, CA: Sage Publications.

Ransdell, S., Arecco, M., Levy, C. (2001). Bilingual longterm working memory: The effects of working memory loads on writing quality and fluency. *Applied Psycholinguistics*, 22, 113-128.

Rescorla, R, Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black & W. Prokasy (Eds.), *Classical conditioning II: Current theory and research* (pp. 64-99). New York: Appleton-Century-Crofts.

Robinson, P., Ellis, N. (Eds.). (2008). *Handbook of cognitive linguistics and second language acquisition.* New York: Routledge.

Romaine, S. (1989). *Bilingualism*. Oxford: Basil Blackwell.

Rommers, J., Meyer, A., Huettig, F. (2013) Object shape and orientation do not routinely influence performance during language processing. *Psychological science*, 24, 2218-2225.

Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4, 328-250.

Rouder, J., Speckman, P., Sun, D., Morey, R., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237.

Sapir, E. (1929). The Status of Linguistics as a Science. In E. Sapir (1958): *Culture, Language and Personality* (ed. D. G. Mandelbaum). Berkeley, CA: University of California Press.

Sato, M., Schafer, A., Bergen, B. (2013). One word at a time: Mental representations of object shape change incrementally during sentence processing. *Language and Cognition,* 5(4), 345-373.

Savickiene, I., Dressler, W. (Eds.) (2007). *Acquisition of diminutives: A cross-linguistic perspective*. Amsterdam: John Benjamins Publishing Company.

Scarcella, R., Oxford, R. (1992). *The Tapestry of Language Learning: The Individual in the Communicative Classroom.* Boston: Heinle.

Sera, M., Berge, C., Del Castillo Pintado, J. (1994). Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive Development*, 9(3), 261-292.

Schacter, D. (2011). *Psychology*. Worth Publishers.

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6, 461-464.

Sjoholm, K. (1995). *The influence of crosslinguistic, semantic, and input factors on the acquisition of English phrasal verbs: A comparison between Finnish and Swedish learners at an intermediate and advanced level.* Abo, Finland: Abo Akademi, Publications of the Department of English.

Sloan Foundation. (1978). *Cognitive Science.* Report of The State of the Art Committee to The Advisors of The Alfred P. Sloan Foundation.

Slobin, D.I. 1996. From "thought and language" to "thinking for speaking." In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity. Studies in the social and cultural foundations of language, vol. 17* (pp. 70-96). Cambridge: Cambridge University Press.

Slobin, D. (2003). Language and thought online. Cognitive consequence of linguistic relativity. In: D. Gentner & S. Goldin-Meadow (Eds.). *Language in mind: Advances in the study of language and thought.* (pp. 157-192). Cambridge, MA: MIT Press.

Spada, N., Lightbown. P. (1993). Instruction and the development of questions in L2 classrooms. *Studies in Second Language Acquisition*, 15, 205-224.

Stanfield, R.A., Zwaan, R. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science,* 12, 153-156.

Sczesny, S., Formanowicz, M., Moser, F. (2016). Can gender-fair language reduce gender stereotyping and discrimination? *Frontiers in Psychology,* 7, Article 25.

Talmy, L. (1975). Semantics and syntax of motion. In: John Kimball (ed.), *Syntax and semantics, 4*, (pp. 181-238), New York: Academic Press.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description, vol. 3: Grammatical categories and the lexicon* (pp. 36-149), Cambridge: Cambridge University Press.

Talmy, L. (2000). Toward a cognitive semantics: Typology and process in concept structuring (Vol. II), MA: MIT Press.

Talmy, L. (2008). Aspects of attention in language. In Robinson, P. & Ellis, N. (Eds.) Handbook of Cognitive Linguistics and Second Language Acquisition. (pp. 27-38). New York, Abingdon: Routledge.

Thomas, M., Van Heuven, W. (2005). Computational models of bilingual comprehension. In: Kroll, J. & De Groot, J. (Eds.), *Handbook of Bilingualism. Psycholinguistic Approaches*. (pp. 202-225). Oxford University Press.

Thompson, A. (2015). Are your participants multilingual? The role of self-assessment in SLA research. Language in Focus Journal, 1(1), 51-65.

Tokowiz, N., Perfetti, C. (2005). Introduction to Part II: Comprehension. In: Kroll, J., De Groot, A. (Eds.) (2005). *Handbook of Bilingualism. Psycholinguistic Approaches* (pp. 173-178). Oxford University Press.

Terrell, T. (1991). The role of grammar instruction in a communicative approach. *Modern Language Journal*, 75, 52-63.

Tomasello, M. (2003). *Constructing a language. A usage-based theory of language acquisition*. Cambridge, Massachusetts, London: Harvard University Press.

Tomczak, E., Ewert, A. (2015). Real and Fictive Motion Processing in Polish L2 Users of English and Monolinguals: Evidence for Different Conceptual Representations. *The Modern Language Journal*, 99, 49-65.

Treffers-Daller, J., Tidball, F. (2015) Can L2 learners learn new ways to conceptualize events? Evidence from motion event construal among English-speaking learners of French.In: Guijarro-Fuentes, P., Schmitz, K. and Müller, N. (eds.) *The acquisition of French in multi-lingual contexts*.

Multilingual Matters, Bristol, pp. 145-184.

Trueswell, J., Papafragou, A. (2010). Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language*, 63, 64–82.

Van Bergen, G., Flecken, M. (2017). Putting things in new places: Linguistic experience modulates the predictive power of placement verb semantics. Journal of Memory and Language, 92, 26-42.

VanPatten, B., Cadierno, T. (1993). Input processing and second language acquisition: A Role for Instruction. *The Modern Language Journal*, 77, 1, 45-57.

VanPatten, B. Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, 15, 225-243.

VanPatten, B., Williams, J., Rott, S. (Eds.), (2004) Form-meaning connections in second language acquisition. Lawrence Erlbaum Associates Publishers, London, Mahwah: New Jersey.

Veivo, O., Jarvikivi, J. (2013). Proficiency modulates early orthographic and phonological processing in L2 spoken word recognition. *Bilingualism: Language & Cognition,* 16(4), 864-883.

Vukovic, N., Williams, J. (2014). Automatic perceptual simulation of first language meanings during second language sentence processing in bilinguals. *Acta Psychologica*, 145, 98-103.

Vukovic, N., Williams, J. (2015). Individual differences in spatial cognition influence mental simulation of language. *Cognition*, 142, 110-122.

Weinreich, U. (1953). *Languages in contact.* The Hague: Mouton.

Whorf, B., Bissel Carroll, J. (1956). *Language, Thought and Reality. Selected Writing of Benjamin Lee Whorf*. MIT Press.

Yin, R. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81, 141-145.

Zwaan, R. (2014). Embodiment and language comprehension: reframing the discussion. *Trends in Cognitive Sciences*.

Zwaan, R. (2014). Replications should be performed with power and precision: A response to Rommers, Meyer, and Huettig (2013). *Psychological Science*, 25, 305-307.

Zwaan, R., Pecher, D. (2012). Revisiting mental simulation in language comprehension: six replication attempts. *PloS One,* 7(12).

Zwaan, R., Stanfield, R., Yaxley, R. (2002). Language comprehenders mentally represent the shape of objects. *Psychological Science*, 13, 168-171.

Appendices

Appendix 1: Language History Questionnaire German

FRAGEBOGEN ZU SPRACHKOMPETENZ

Es folgen einige Fragen zu ihrem Bildungsstand und Sprachgebrauch. Bitte beantworte diese Fragen so vollständig wie möglich.

PERSÖNLICHER HINTERGRUND:

Alter:
Geschlecht:

Welche ist ihre dominante Hand? Links / Rechts
Ich habe eine normale Sehfähigkeit: Ja / Nein

Was studieren Sie? oder was ist Ihr Beruf?
In welchem Institut lernen Sie Spanisch?
Auf welchem Niveau lernen Sie Spanisch? A2, B1, B2, C1, C2

Sind Sie in Deutschland geboren worden? Ja / Nein
Falls ja:
Haben Sie seit ihrer Geburt immer in Deutschland gelebt? Ja / Nein

Falls nein:
Wo haben Sie sonst gelebt und wie lange?
Wie alt waren Sie, als Sie nach Deutschland kamen?
Wie lange leben Sie bereits in Deutschland?
Sind Sie zwischenzeitig in ihr Geburtsland zurückgekehrt für mehr als 6 Monate (falls ja, wie lange?)? Ja / Nein

SPRACHLICHER HINTERGRUND:

Was ist ihre Muttersprache?

Bitte listen Sie alle anderen Sprachen auf, die Sie beherrschen. Bewerte jeweils, wie gut Sie die

Sprache beherrschen nach folgender Skala:

Nicht gut 1 2 3 4 5 Sehr gut

Sprache	Sprechen	Hörverstehen	Schreiben	Lese-verstehen	Grammatik	Aussprache
1 Spanisch						
2						
3						
4						
5						

Bitte geben Sie für jede oben aufgelistete Sprache an, wo und in welchem Alter Sie sie gelernt haben sowie, falls zutreffend, ob Sie sie mit formeller Anleitung oder in informellem Kontext gelernt haben.

Sprache	Land	Alter	Unterrischt (ja/nein)	Dauer des Unterrichts	Informell (ja/nein)	Dauer des informellen Lernens
Spanisch						

Bewerte für jede angegebene Sprache, wie sehr Sie den folgenden Aussagen zustimmen nach folgender Skala.

Stimme nicht zu 1 2 3 4 5 Stimme zu

Sprache	Ich spreche diese Sprache gerne	Ich fühle mich sicher in dieser Sprache	Ich denke, es ist wichtig, diese Sprache gut zu können
Spanisch			

Welche der oben angeführten Sprachen sprechen Sie mit welchen der folgenden Personen(gruppen), wie viele Stunden am Tag, für welche Themen und an welchem Ort (zu Hause, bei der Arbeit, etc.)

	Sprache	Stunden am Tag	Thema	Ort
Mutter				
Vater				
ältere(r) Bruder / Schwester				
jüngere(r) Bruder / Schwester				
Kinder				
andere Familienmitglieder				
Mitbewohner				
Partner				
Freunde				
Kollegen				

Welche der angegebenen Sprachen benutzen Sie für die folgenden Aktivitäten und wie viele Stunden pro Tag?

Aktivität	Sprache	Stunden pro Tag
Lesen		
Fernsehen		
Radio hören		
Email, Internet		
Spielen		

Im Allgemeinen, wie gerne lernen Sie neue Sprachen?

Nicht gerne 1 2 3 4 5 Gerne

Im Allgemeinen, wie leicht fällt es Ihnen, neue Sprachen zu lernen?

Schwer 1 2 3 4 5 Leicht

Falls Sie noch andere Anmerkungen zu ihrem sprachlichen Hintergrund haben, die Sie für wichtig für ihre Fähigkeit, diese Sprachen zu benutzen, halten, schreibe diese gerne hier auf:

……
……
……

Welche Lehrbücher verwenden Sie oder haben Sie verwendet um Spanisch zu lernen?

__ Con gusto, A2 __ Gente 2, B1

__ E*ñe, B1-B2* __ *Gent*e 3, B2

__ Gente hoy 1, A1-A2 __ *Via rápida, A1,* A2, B1+

__ Gente hoy 2, B1-B2 __ ……..

Wussten Sie schon vor dem Test heute dass es Augmentative gibt? Ja / Nein

Falls ja:
Wie lange kennen Sie die Augmentative schon?
Wie oft stoßen Sie auf oder bemerken Sie Augmentative wenn Sie Spanisch lesen oder hören?
Sehr oft Oft Gelegentlich Selten Nie

Wie oft benutzen Sie Augmentative wenn Sie Spanisch schreiben oder sprechen?
Sehr oft Oft Gelegentlich Selten Nie

Appendix 2: Language History Questionnaire Spanish

CUESTIONARIO DE CONOCIMIENTO DE IDIOMAS

A continuación se presentan algunas preguntas sobre su educación y uso del lenguaje. Por favor, conteste a estas preguntas de la manera más completa posible.

ANTECEDENTES:

Nombre, apellido:
Edad:
Sexo:

Mano dominante: izquierda / derecha
Tengo visión normal: Sí / No

¿Qué carrera/grado estudia?
¿En cual Instituto estudias alemán?
¿Cual nivel de alemán tienes? A2/B1/B2/C1/C2

¿Ha nacido en España? Sí / No

En caso afirmativo:
¿Ha vivido en España desde que nació? Sí / No

Si no:
¿En qué otros lugares ha vivido usted?
¿Qué edad tenía cuando vino a España?
¿Cuánto tiempo ha estado viviendo en España?
¿Ha regresado a su país de nacimiento durante más de 6 meses (en caso afirmativo, por cuánto tiempo?) Sí / No

CONOCIMIENTOS DE IDIOMAS

¿Cuál es su lengua materna?

Por favor mencione los idiomas que conoce. Para cada uno, indique su conocimiento en las diferentes áreas, utilizando la siguiente escala:

No es bueno 1 2 3 4 5 Muy bueno

Idioma	Expresión oral	Comprensión oral	Expresión escrita	Comprensión lectora	Gramática	Pronunciación
1 Aleman						
2						
3						
4						
5						

Para cada idioma que ha mencionado en el punto anterior, indique por favor el lugar y la edad en que los aprendió, y si los aprendió en un contexto formal de clase o en un contexto informal (por ejemplo, hablando con los hablantes nativos del idioma)

Idioma	País	Edad	Clases(sí/no)	Duración de las clases	Aprendizaje informal (sí/no)	Duración del aprendizaje informal
Aleman						

Para cada uno de los idiomas que ha mencionado, indique por favor si está de acuerdo con estas afirmaciones, utilizando la siguiente escala

Desacuerdo 1 2 3 4 5 De Acuerdo

Idioma	Me gusta hablar este idioma	Me siento seguro/a usando este idioma	Creo que es importante tener un buen conocimiento de este idioma
Aleman			

Para cada uno de los idiomas que ha mencionado, indique por favor cuál utiliza con las siguientes

personas, durante cuántas horas al día, sobre qué temas habla (familia, estudios, etc.) y el lugar dónde lo utiliza (casa, trabajo, etc.):

	Idioma	Horas al día	Tema	Lugar
Madre				
Padre				
Hermano/Hermana mayor				
Hermano/Hermana menor				
Niños				
Otros miembros de la familia				
Compañeros de piso				
Pareja				
Amigos				
Colegas				

Para cada uno de los idiomas que ha mencionado, indique por favor qué idioma utiliza para las siguientes actividades y por cuántas horas al día

Actividad	Idioma	Horas al día
Lectura		
Ver la TV		
Escuchar la radio		
Correo electrónico, Internet		
Juegos		

En general, ¿hasta qué punto le gusta aprender nuevos idiomas?

No me gusta 1 2 3 4 5 Me gusta

En general, ¿hasta qué punto le resulta difícil o fácil aprender nuevos idiomas?

Difícil 1 2 3 4 5 Fácil

Si tiene algún comentario sobre los idiomas que conoce y piensa que es importante para el uso que hace de los mismos, por favor menciónelos aquí:

……
……
……

¿Cuáles libros de texto utiliza o ha utilizado para aprender alemán?

__ Menschen A1,A2 __ Aspekte 3

__ Menschen B1 __ Erkundungen

__ Aspekte B1+ __ ……

__ Ziel B2

¿Sabía antes de la prueba hoy de los verbos *legen/stellen/liegen/stehen*? Sí / No

Si, sí:
¿Cuánto tiempo hace que conoce estos verbos?
¿Con cuál frecuencia se encuentra o observa estos verbos leyendo o escuchando alemán?

Con mucha frecuencia A menudo De vez en cuando Rara vez Nunca

¿Con qué frecuencia se utilizan estos verbos cuando escribe o habla alemán?

Con mucha frecuencia A menudo De vez en cuando Rara vez Nunca

Appendix 3: Description of CEFR Proficiency Levels

EUROPEAN LEVELS - SELF ASSESSMENT GRID

		A1	A2	B1	B2	C1	C2
UNDERSTANDING	Listening	I can understand familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.	I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements.	I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.	I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programmes. I can understand the majority of films in standard dialect.	I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. I can understand television programmes and films without too much effort.	I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided. I have some time to get familiar with the accent.
	Reading	I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues.	I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.	I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.	I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.	I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field.	I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.
SPEAKING	Spoken Interaction	I can interact in a simple way provided the other person is prepared to repeat or rephrase things at a slower rate of speech and help me formulate what I'm trying to say. I can ask and answer simple questions in areas of immediate need or on very familiar topics.	I can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar topics and activities. I can handle very short social exchanges, even though I can't usually understand enough to keep the conversation going myself.	I can deal with most situations likely to arise whilst travelling in an area where the language is spoken. I can enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).	I can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible. I can take an active part in discussion in familiar contexts, accounting for and sustaining my views.	I can express myself fluently and spontaneously without much obvious searching for expressions. I can use language flexibly and effectively for social and professional purposes. I can formulate ideas and opinions with precision and relate my contribution skilfully to those of other speakers.	I can take part effortlessly in any conversation or discussion and have a good familiarity with idiomatic expressions and colloquialisms. I can express myself fluently and convey finer shades of meaning precisely. If I do have a problem I can backtrack and restructure around the difficulty so smoothly that other people are hardly aware of it.
	Spoken Production	I can use simple phrases and sentences to describe where I live and people I know.	I can use a series of phrases and sentences to describe in simple terms my family and other people, living conditions, my educational background and my present or most recent job.	I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or relate the plot of a book or film and describe my reactions.	I can present clear, detailed descriptions on a wide range of subjects related to my field of interest. I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.	I can present clear, detailed descriptions of complex subjects integrating sub-themes, developing particular points and rounding off with an appropriate conclusion.	I can present a clear, smoothly-flowing description or argument in a style appropriate to the context and with an effective logical structure which helps the recipient to notice and remember significant points.
WRITING	Writing	I can write a short, simple postcard, for example sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form.	I can write short, simple notes and messages. I can write a very simple personal letter, for example thanking someone for something.	I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions.	I can write clear, detailed text on a wide range of subjects related to my interests. I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view. I can write letters highlighting the personal significance of events and experiences.	I can express myself in clear, well-structured text, expressing points of view at some length. I can write about complex subjects in a letter, an essay or a report, underlining what I consider to be the salient issues. I can select a style appropriate to the reader in mind.	I can write clear, smoothly-flowing text in an appropriate style. I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works.

Appendix 4: Instruction on the German verbs *legen/stellen*, including 18-item posttest

<u>Die Verben *stellen, legen, liegen* und *stehen*</u>

Wussten Sie schon?
Auf Deutsch gibt es Verben, die die Position von Objekten oder Personen beschreiben. Diese Verben zeigen an ob das Objekt oder die Person sich in einer vertikalen oder horizontalen Position befindet. Vier Verben, die oft benutzt werden, sind:

1. stellen poner en una posición vertical
2. legen poner en una posición horizontal
3. liegen estár en una posición horizontal
4. stehen estár en una posición vertical

Falls man die Position oder Bewegung noch genauer beschreiben möchte, kann man im Deutschen noch *gerade-* oder *hin-* hinzufügen (zum Beispiel: *geradestellen, geradelegen; hinstellen, hinlegen*).

Konjugationen

	stellen	**stehen**	**legen**	**liegen**
Ich	stelle	stehe	lege	liege
Du	stellst	stehst	legst	liegst
Er/sie	stellt	steht	legt	liegt
Wir	stellen	stehen	legen	liegen
Ihr	stellt	steht	legt	liegt
Sie	stellen	stehen	legen	liegen

Beispiel

Ulrich: "Uf, dieses Buch ist sehr groß und schwer!"

Kathi: "Du Armer! Komm, dann **legen** wir das auf den Tisch!"

Übung

Geben Sie an, ob die folgenden Verben eine horizontale, vertikale oder neutrale Position bezeichnen. Zeichnen Sie bitte einen Kreis um die richtige Antwort:

19. beantragen	horizontal / vertikal / neutral
20. fordern	horizontal / vertikal / neutral
21. geradelegen	horizontal / vertikal / neutral
22. liegen	horizontal / vertikal / neutral
23. zuweisen	horizontal / vertikal / neutral
24. legen	horizontal / vertikal / neutral
25. ergeben	horizontal / vertikal / neutral
26. stehen	horizontal / vertikal / neutral
27. hintereinanderstellen	horizontal / vertikal / neutral
28. geradestellen	horizontal / vertikal / neutral
29. wegstellen	horizontal / vertikal / neutral
30. weglegen	horizontal / vertikal / neutral
31. platzieren	horizontal / vertikal / neutral
32. stellen	horizontal / vertikal / neutral
33. hinstellen	horizontal / vertikal / neutral
34. hinlegen	horizontal / vertikal / neutral
35. machen	horizontal / vertikal / neutral
36. aufeinanderlegen	horizontal / vertikal / neutral

Appendix 5: Instruction on Spanish augmentatives, including 18-item posttest

<u>Los aumentativos</u>

Sabías que:
En español, además de sufijos diminutivos (cas**ita**) hay unos sufijos llamados aumentativos que se pueden añadir a sustantivos y adjetivos, y su significado es expresar tamaño grande. Así podemos decir que algo es grande sin tener que añadir un adjetivo como *grande* or *enorme.*

En español los aumentativos más frecuentes son:

masculino **-ón** **-azo** **-ote**

feminino **-ona** **-aza** **-ota**

-ON/-ONA	**hombre**	Mann	hombr**ón**	großer Mann
	carpeta	Mappe	carpet**ona**	große Mappe
-AZO/-AZA	**perro**	Hund	perr**azo**	großer Hund
	botella	Flasche	botell**aza**	große Flasche
-OTE/-OTA	**libro**	Buch	libro**te**	großes, schweres Buch
	fruta	Frucht	frut**ota**	große Frucht

Uso de los aumentativos

1. Las palabras que terminan en consonante: añaden el sufijo completo:
Mujer Frau mujer**ona** große Frau

2. Las palabras que terminan en vocal: la vocal se quita antes de añadir el sufijo:
carpeta Mappe **carpetona** große Mappe

Ejemplo:

Juan: "Ya has leído *El Quijote* de Cervantes? Es un poco gordo, pero muy interesante.

María: "¡Uf! Estos **librotes** son muy difíciles de leer!"

Ejercicio

Indica si los sustantivos que aparecen a continuación expresan tamaño pequeño, normal o grande. Pon un círculo alrededor de la respuesta correcta:

1. bolsita pequeño / normal / grande
2. teléfono pequeño / normal / grande
3. manzana pequeño / normal / grande
4. puertita pequeño / normal / grande
5. maletona pequeño / normal / grande
6. papelito pequeño / normal / grande
7. niño pequeño / normal / grande
8. plátano pequeño / normal / grande
9. ventanita pequeño / normal / grande
10. silla pequeño / normal / grande
11. zapatón pequeño / normal / grande
12. casota pequeño / normal / grande
13. gatote pequeño / normal / grande
14. pelotaza pequeño / normal / grande
15. vasito pequeño / normal / grande
16. platazo pequeño / normal / grande
17. lápiz pequeño / normal / grande
18. perrito pequeño / normal / grande

Appendix 6: Wordlist L2 German with L1 Spanish translations

Vokabeln

in die Experimente werdest du eine Serie deutsche Worte sehen. Es ist wichtig dass du verstehest was diese Worten bedeuten. Hierunter findest du eine Wortliste mit spanischer Übersetzung. Bitte, benutz 5 Minuten um diese Worten zu lehren. Wenn du bereit bist, kannst du anfangen mit dem ersten Experiment.

1.	die Treppe	la escalera
2.	der Ball	la pelota
3.	das Lineal	la regla
4.	die Tomate	el tomate
5.	das Schneidebrett	la tabla de cortar
6.	die Taschenlampe	la linterna
7.	das Papier	el papel
8.	die Trompete	la trompeta
9.	die CD	el CD
10.	die Tube	el tubo
11.	der Stuhl	la silla
12.	der Klebestift	la barra de pegamento
13.	die Säge	la sierra
14.	das Deodorant	el desodorante
15.	die Waschmaschine	la lavadora
16.	die Batterie	la pila
17.	die Kerze	la vela
18.	der Stern	la estrella
19.	das Fernglas	el par de binoculares
20.	das Regal	el estante
21.	das Handtuch	la toalla
22.	der Lippenstift	la barra de labios
23.	die Pfanne	la sartén
24.	der Tisch	la mesa
25.	der Stift	el lápiz
26.	die Glocke	la campana
27.	die Bandspule	el carrete del Hilo
28.	das Brett	el plato
29.	der Schrank	el armario
30.	die Matte	la estera

Appendix 7: Wordlist L2 Spanish with L1 German translations

Lista de palabras

En los experimentos vas a ver una serie de palabras en español. Es importante que conozcas el significado de estas palabras. Abajo tienes la lista de palabras con su traducción al alemán. Por favor usa 5 minutos para estudiar las palabras. Cuando sepas el significado de todas las palabras, puedes comenzar con el primer experimento.

1.	la escalera	die Treppe
2.	la pelota	der Ball
3.	la regla	das Lineal
4.	el tomate	die Tomate
5.	la tabla de cortar	das Schneidebrett
6.	la linterna	die Taschenlampe
7.	el papel	das Papier
8.	la trompeta	die Trompete
9.	el CD	die CD
10.	el tubo	die Tube
11.	la silla	der Stuhl
12.	la barra de pegamento	der Klebestift
13.	la sierra	die Säge
14.	el desodorante	das Deodorant
15.	la lavadora	die Waschmaschine
16.	la pila	die Batterie
17.	la vela	die Kerze
18.	la estrella	der Stern
19.	el par de binoculares	das Fernglas
20.	el estante	das Regal
21.	la toalla	das Handtuch
22.	la barra de labios	der Lippenstift
23.	la sartén	die Pfanne
24.	la mesa	der Tisch
25.	el lápiz	der Stift
26.	la campana	die Glocke
27.	el carrete del hilo	die Bandspule
28.	el plato	das Brett
29.	el armario	der Schrank
30.	la estera	die Matte