



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

CROSS-LINGUISTIC METAPHOR INTELLIGIBILITY BETWEEN ENGLISH AND GERMAN

Christoph Hesse



A thesis submitted in fulfilment of requirements for the degree of

DOCTOR OF PHILOSOPHY

to Linguistics and English Language

School of Philosophy, Psychology and Language Sciences

UNIVERSITY *of* EDINBURGH

September 2015

Supervisors: Prof Ronnie Cann and Dr Hannah Rohde

External examiner: Dr John Tomlinson

Internal examiner: Dr Chris Cummins

Copyright © Christoph Hesse 2015

www.christophhesse.com

Christoph Hesse asserts the moral right to
be identified as the author of this work.

A catalogue record of this work is available
from Edinburgh University Library.

Set in Helvetica, Arial, Times New Roman,
and the Computer Modern font family.

Typeset in L^AT_EX developed by Leslie Lamport.

Printed and bound in Edinburgh in 2015.

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author unless otherwise stated. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis may not be reproduced or quoted extensively from without first obtaining permission in writing from the author. The content must not be changed in any way or published or sold commercially in any format or medium, electronic, mechanical, photocopying, recording or otherwise, without the prior formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Declaration

I hereby declare that this thesis is of my own composition and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Christoph Hesse

Abstract

Conceptual Metaphor Theory (CMT, Lakoff & Johnson 1980; Lakoff, 1983, 1987, 1993, 2008, 2009), the most prominent cognitive approach to metaphor comprehension, argues that the nature of interconnections within the conceptual system is inherently metaphoric-analogical and that systematic patterns in linguistic metaphor reveal these cognitive interconnections. Relevance Theory (RT, Sperber & Wilson, 1986; Wilson & Sperber, 1993; Sperber & Wilson, 1995; Wilson & Sperber, 2002, 2004) and Graded Salience (GS, Giora, 1997, 1998, 1999, 2003; Peleg et al., 2008; Peleg & Giora, 2011) disagree that systematic patterns in linguistic metaphor can be taken as direct evidence of their cognitive representation.

A metaphor consists of two concepts, a source and a target concept. The metaphor implies an analogy between the two concepts. To comprehend a metaphor is to infer under which conditions the implied analogy holds. The meaning of the two concepts is pragmatically enriched by these additional assumptions. Metaphor comprehension is an inferential process. The result of this process is the enriched meaning of the metaphor. This meaning can become conventionalised, in which case it often serves as an inferential shortcut: instead of having to consider all conceptually possible interpretations and their plausibility in the context of the analogy, speakers who are familiar with the conventional (i.e. idiomatic) meaning are provided with a default interpretation.

According to CMT, the inferential process is a process of interconnecting primary embodied concepts to ever more complex higher-order concepts. On this view, a metaphoric idiomatic meaning is such a complex concept where the conceptual interconnections are conventional. According to RT, the inferential process is a process of inferring a meaning that is in line with the speaker's communicative intent, the discourse context, and interlocutors' expectations of the cognitive relevance of potential inferences. On this view, metaphoric idiomatic meanings are highly salient inferences with a high degree of contextual relevance because speakers' expectations of relevance are conventionalised. According to GS, the inferential process consists of two modules that work in parallel: a module that infers salient meanings based on linguistic knowledge and a module that enriches the meaning by taking non-linguistic knowledge such as conceptual, experiential, perceptual, contextual, and world knowledge into consideration. On this view, metaphoric idiomatic meanings are highly salient inferences because of speakers' knowledge of non-conceptual linguistic conventions.

This thesis investigates the claims made by CMT, RT and GS by experimentally testing the cross-linguistic communicability of metaphoric proverbs with idiomatic meanings. Proverbs are selected such that the similarity of metaphors' source and target concepts, expectations of contextual relevance, and the degree of familiarity with proverbs' conventional wording is cross-linguistically maximised. If CMT is correct, then when cross-linguistic conceptual similarity is maximised in this way, monolingual native speakers should find L2 language-specific metaphors communicable. If RT and GS are correct, then monolingual native speakers should find L2-specific metaphors less communicable than L1-specific and non-language-specific metaphoric proverbs because they lack knowledge of the necessary non-conceptual linguistic conventions. Cross-linguistic metaphor communicability is measured in three ways in the experiments: (1) through reading/response times, (2) through plausibility judgements, and (3) through a context creation task. Results show that cross-linguistic metaphor communicability of L2-specific metaphors is lowered for monolingual native speakers on all three measures.

Acknowledgements

My heartfelt thank you goes to the following people. My amazing, helpful, supportive, patient, and enthusiastic supervisors Prof Ronnie Cann and Dr Hannah Rohde. My examiners Dr John Tomlinson and Dr Chris Cummins for their excellent and valuable comments on my thesis and for making the examination both productive and enjoyable. Prof Regine Eckardt who introduced me to Gricean pragmatics. Prof Anke Holler who taught me psycholinguistics. Dr Manfred Sailer who taught me second-order predicate calculus, lambda calculus, and introduced me to Frege's work. Reinhard Baukowitz who introduced me to computational linguistics. All of my other teachers. The Syntax & Semantics Reading Group and fellow PhD students at Edinburgh University.

Hilko Hoffmann, Kristian Sons, Boris Brönnner, Léa Yvonne Schäfer, Prof Philipp Slusallek and the rest of the team at the Deutsches Forschungszentrum für Künstliche Intelligenz (German Research Centre for Artificial Intelligence), Saarbrücken and Kaiserslautern.

My family, Barbara, Lilo, and Martin for their unwavering strength and support. The best friends anyone could have, Stefan John and family, Sven Murlowsky and family, Sascha & Anke Bargmann, Carla Xavier & Luca Rognoni, Brenna Barks, Maria Polyakova, Nina Hölscher and family, Liam Feder, Stefanie Röhrig, Mario Linz, Christine Maier, Rachael Bailes & Justin Quillinan, Jie Yang, Hannah Cornish, Sean Roberts, Bill Thompson, Amanda Cardoso, Vanessa Ferdinand, Chrissy Cuskley and family, Juan Lewis and family, Laura Sterian and family, Gloria Chamorro, Matt Spike and family, Kevin Stadler, James Winters, Thijs Lubbers, Catriona Silvey, Ashley Taylor-Richards, John-Sebastian Schutter. My former and best flatmates ever, Luis Pineda, Merlin Lao, Wallace Ye, and Timur Yusupov.

The attendees of the 4th International Summer School in Cognitive Science and Semantics 2012, in particular Deirdre Wilson, Dan Sperber, Ernest Lepore, Peter Ludlow, Victor Fernández Castro, Stefano Di Piazza, Dan Zeman, Christopher Millano, Sandy DeVries, and Anna Drożdżowicz. The audience at the Linguistics & English Language Postgraduate Conferences 2011, 2012, and 2013 for their many helpful questions and comments. The audience at the 36th, 37th, and 38th Language at Edinburgh Lunches. The audience at AMLaP 2013 in Marseille, in particular Prof Laurie Feldman. The audience at the 1st University College London Graduate Conference in Linguistics 2013, in particular Deirdre Wilson, Robyn Carston, Rachel Giora, Diana Mазzarella, Sophie Schieber, and Patrick Elliott. The audience at the Linguistics & English Language lecture in the spring of 2012 and my tutees in the years 2010 through 2014.

The participants in my experiments, undergraduate and postgraduate students at Edinburgh University, Göttingen University, Saarbrücken University, the University of Mainz, Frankfurt University, and in Arnstadt, Erfurt, Dannheim, and Molsdorf, Thüringen.

Contents

Acknowledgements	5
List of Figures	10
List of Tables	11
1 Introduction	13
1.1 Theories of metaphor comprehension	13
1.2 The experimental rationale	17
1.3 Thesis overview	20
2 Concepts	22
2.1 Introduction	22
2.2 Concepts as mental representations	23
2.3 Embodied Cognition and Conceptual Metaphor Theory	29
2.3.1 Inferential comprehension as creating ever more complex higher-order concepts from primary embodied concepts	29
2.3.2 Metaphor systematicity as evidence for the structure of the conceptual system and of meaning salience	33
2.4 Concepts as abilities	37
2.5 Concepts as Fregean senses	39
2.6 Linguistic meaning	44
2.7 Concepts from a cross-linguistic perspective	47
2.8 Summary	54
3 Metaphor	56
3.1 Introduction	56
3.2 The metaphor comprehension process	57
3.2.1 The implied analogy as a failure of direct attribution	58
3.2.2 The implied analogy as indirect attribution via <i>ad hoc</i> concepts and prag- matic enrichment	61
3.2.3 The implied analogy as a complex concept composed from embodied pri- mary concepts through conceptual metaphors	63
3.2.4 Inferential strength	64
3.2.5 Meaning salience	65
3.3 Metaphors with conventional meanings	68

3.3.1	Idiomatic meanings as inferential shortcuts	70
3.3.2	A general theory of inferential comprehension	71
3.3.3	Toward testing RT, GS, and CMT's predictions concerning metaphoric idiomatic meanings	75
3.4	Cross-linguistic metaphor intelligibility	79
3.4.1	Cross-linguistic approximate conceptual similarity and cross-linguistic metaphor intelligibility	79
3.4.2	Cross-linguistic correlation and non-correlation between the linguistic form and idiomatic meanings of proverbs	80
3.5	Summary	84
4	Methodology: cross-linguistic metaphor intelligibility	87
4.1	Introduction	87
4.2	Research questions	87
4.3	Toward a methodology to maximise the chance of cross-linguistic metaphor intelligibility	90
4.3.1	Splitting metaphor conventionality into familiarity with proverbs' linguistic form and the arbitrariness of the form-meaning association	94
4.3.2	Corpus frequency as an index of familiarity with proverbs' form	95
4.3.3	Conceptual plausibility as an index of the arbitrariness of the form-meaning association	95
4.4	Definitions and operationalisations	98
4.4.1	Cross-linguistic intelligibility	98
4.4.2	Speaker groups	99
4.4.3	Measuring cross-linguistic metaphor intelligibility	100
4.4.4	Closeness of language contact between English and German	104
4.4.5	Language specificity of linguistic metaphors	105
4.4.6	Context and fillers	111
4.5	Summary	112
5	Material: cross-linguistic comparability of metaphors' linguistic form	114
5.1	Introduction	114
5.2	Metaphor selection process	116
5.3	Corpus frequency as an indicator of speakers' familiarity with metaphors' linguistic form	117
5.4	Frequency synchronic: The frequency distribution of an entire vocabulary	121
5.4.1	Cross-linguistic comparability of metaphors' corpus frequency and familiarity probabilities	129
5.4.2	Lexical frequency of metaphors' source and target concepts	131
5.5	Frequency diachronic: stability over time	132
5.6	How do we know zero frequency means zero?	134
5.7	Substantive and statistical significance of differences in corpus frequency	137
5.8	Cross-linguistic comparability of metaphors' length	144
5.9	Summary	151

6	Material: cross-linguistic comparability of metaphors' linguistic meaning	153
6.1	Introduction	153
6.2	The lexical semantics of metaphors' source and target concepts	154
6.3	Metaphors' compositional semantics and idiomatic meanings	157
6.4	Summary	184
7	Experiments: cross-linguistic metaphor intelligibility out of context	186
7.1	Introduction	186
7.2	Predictions	187
7.3	Material and methods	188
7.3.1	Participants	188
7.3.2	Material	189
7.3.3	Procedure	190
7.3.4	Design	192
7.4	Results	192
7.4.1	Reading/response times	193
7.4.2	Plausibility judgements	205
7.4.3	Metaphor frequency as covariate	212
7.4.4	The context creation task	213
7.5	Summary	231
8	Experiments: cross-linguistic metaphor intelligibility in context	233
8.1	Introduction	233
8.2	Predictions	234
8.3	Definitions and operationalisations	234
8.4	Material and methods	235
8.4.1	Participants	235
8.4.2	Metaphor material	236
8.4.3	Context material	236
8.4.4	Procedure	249
8.4.5	Design	251
8.5	Results	251
8.6	The reliability of the experimental results	260
8.7	Summary	262
9	Summary and conclusions	265
9.1	Introduction	265
9.2	Theoretical background	265
9.3	The experimental rationale	269
9.4	The degree to which maximising cross-linguistic approximate conceptual similarity has been achieved	272
9.5	Experimental findings	275
9.6	General conclusion	281
	Bibliography	282

A Appendix	294
A.1 Experimental material	294
A.2 Corpus frequencies	296
A.3 Semantic Levenshtein distances	303
A.4 Statistical analysis of the experimental results	305
A.5 Experimental results and deficits in the material	309

List of Figures

4.1	Penrose triangle, infinite staircase, and M.C. Escher’s lithograph <i>Ascending and Descending</i>	102
4.2	Book translations: English-German, German-English.	104
5.1	The raw distribution of the number of wordforms per frequency $n(f)$ for the vocabularies of British English, American English, and German.	126
5.2	The probability mass function $f_X(x) = \Pr(X = x)$ for the vocabularies of British English, American English, and German.	128
5.3	The cumulative probability distribution function $F_X(x) = \Pr(X \leq x)$ for the vocabularies of British English, American English, and German.	129
5.4	Diachronic development of <i>silence is golden</i> (Adj) and <i>silence is gold</i> (N) in English between 1810 and 2000.	133
5.5	Histogram for the number of words per full stop.	136
5.6	The number of words per full stop and their standard deviation as a measure of a corpus’ measurement accuracy.	137
5.7	Difference in metaphor length between the English and German version of the material.	146
7.1	Reading/response times for English monolinguals and English-German bilinguals.	193
7.2	Difference in reading/response times.	195
7.3	Metaphor plausibility judgements: out-of-context condition.	206
8.1	Metaphor plausibility judgements: out-of-context and in-context condition.	252
A.1	Diachronic change in frequency: English-specific metaphors.	300
A.2	Diachronic change in frequency: German-specific metaphors.	301
A.3	Diachronic change in frequency: culturally-shared metaphors.	302
A.4	Residual plots for reading/response times.	305
A.5	Residual probability density histograms for reading/response times.	306
A.6	Residual plots for out-of-context and in-context plausibility judgements.	307
A.7	Residual probability density histograms for out-of-context and in-context plausibility judgements.	308

List of Tables

2.1	Eight types of linguistic symbols based on the three binary features <i>concept</i> , <i>sense relations</i> , and <i>external referent</i>	45
2.2	Cross-linguistic differences in Fregean sense relations as evidence for conceptual differences.	50
3.1	Cross-linguistic differences in metaphoric proverbs as evidence for differences in conceptual metaphors.	81
4.1	Number of participants.	99
4.2	Mean age of participants.	100
4.3	Examples of the different metaphor types.	105
5.1	Number of the different metaphor types.	116
5.2	Average corpus frequencies F and cumulative probabilities p	130
5.3	Corpus-based familiarity probability.	141
5.4	Hypotheses confirmation and statistical tests.	143
6.1	Relative Levenshtein distance of lexical semantics.	157
6.2	Relative Levenshtein distance of compositional semantics.	173
6.3	Cumulative differences between the English and German version of the experimental material in terms of their linguistic form and meaning.	175
6.4	Cross-linguistic variation in the experimental material: English and German metaphors.	180
6.5	Cross-linguistic variation in the experimental material: cross-linguistically shared and novel metaphors.	181
6.6	Cross-linguistic variation in the experimental material: metaphors' features.	182
6.7	Cross-linguistic variation in the experimental material: metaphors' linguistic form and meaning.	182
7.1	Median reading/response times.	195
7.2	Variance explained by each fixed and random effect in the linear mixed effects models: reading/response time models.	200
7.3	Out-of-context reading/response times.	202
7.4	Variance explained by each fixed and random effect in the linear mixed effects models: out-of-context plausibility judgement models.	208
7.5	Out-of-context plausibility judgements.	209

8.1	Variance explained by each fixed and random effect in the linear mixed effects models: in-context plausibility judgement models.	253
8.2	In-context plausibility judgements.	255
8.3	Effect of context on plausibility judgements.	257
8.4	Statistical reliability of the significance of differences in reading/response times.	260
8.5	Statistical reliability of the significance of differences in out-of-context plausibility judgements.	261
8.6	Statistical reliability of the significance of differences in in-context plausibility judgements.	261
9.1	First experimental finding.	276
9.2	Second experimental finding.	280
A.1	Metaphors and continuations used in the experiments.	294
A.2	Word frequencies.	296
A.3	Metaphor frequencies.	298
A.4	Diachronic change in metaphors' corpus frequency.	299
A.5	Relevant and irrelevant sense differences for all translation pairs.	303
A.6	Cross-linguistic variation in the experimental material: English and German metaphors.	309
A.7	Cross-linguistic variation in the experimental material: cross-linguistically shared and novel metaphors.	310
A.8	Cross-linguistic variation in the experimental material: metaphors' features.	311
A.9	Cross-linguistic variation in the experimental material: metaphors' linguistic form and meaning.	311
A.10	Variation in experimental results due to metaphor type.	312
A.11	Variation in experimental results neither explained by metaphor type nor inter-personal variation nor variation between metaphors of the same type.	313

Chapter 1

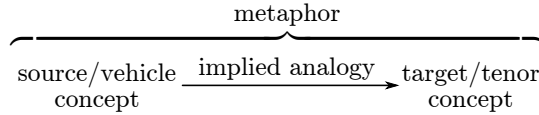
Introduction

1.1 Theories of metaphor comprehension

Meaning salience plays a central role in theories of inferential comprehension. It is reasonable to assume that the mind and thus the inferential process, being a mental process, is parsimonious. The goal of the inferential process is to derive a meaning that is a plausible meaning of an utterance in a given discourse context and to do that by spending the least amount of cognitive effort in doing so. Since this process is geared towards efficiency, preferential treatment is given to drawing those inferences that are most helpful in deriving a satisfactory meaning. The preferential treatment that determines the likelihood with which inferences are drawn and the prominence that they have during processing, assigns a degree of salience to each inference during processing. This inferential likelihood and prominence is known as *meaning salience*. The three theories of inferential processing that I am interested in in my thesis are Conceptual Metaphor Theory (CMT) (Lakoff & Johnson, 1980; Lakoff, 1983, 1987; Lakoff & Johnson, 1999, 2003; Lakoff, 2008, 2009), Relevance Theory (RT) (Sperber & Wilson, 1986; Wilson & Sperber, 1993; Sperber & Wilson, 1995; Wilson & Sperber, 2002, 2004), and the Graded Salience Hypothesis (GS) (Giora, 1997, 1998, 1999, 2003; Peleg, Giora & Fein, 2008; Peleg & Giora, 2011). I am interested in these three in particular because they all take the inferential process and the meanings that are derived to be inherently context-dependent, but they take different perspectives on what determines meaning salience. Conventionality is a main factor in determining meaning salience. Conventional linguistic meanings such as idiomatic meanings have the highest salience. CMT claims that conventionalisation is a conceptual process; RT and GS claim that it is a non-conceptual process. The experiments of this thesis investigate whether conceptual or non-conceptual aspects of the idiomatic meanings of metaphoric proverbs have a greater influence on speakers' inferential comprehension.

A metaphor, such as *time is money*, consists of three things: (1) a 'source' or 'vehicle' concept, MONEY, (2) a 'target' or 'tenor' concept, TIME, and (3) an 'implied analogy' that invites us to comprehend the target concept in terms of the source concept. Historically, there are generally three views on what concepts are: (1) they are mental representations of sensory perception, experience out of interaction with the world, and imagination. (2) Concepts are best understood through our ability to use them. If the successful completion of a specific task requires that we have a particular concept, then a demonstration of our ability to complete the task should be indicative of us having that concept. (3) The view that concepts are Fregean

sense relations, i.e. that they are abstract symbols whose meaning is defined in the propositional system of a language and by conditions of truth in the world. The physiological and behavioural evidence supports the mental representations view most and its modern version has become known as Embodied Cognition. On this view all basic primary concepts in the conceptual system (speakers' mental lexicons) are rooted in perceptual experience or introspective sensory experience of body-internal states such as emotions.



To comprehend a metaphor is to infer under which conditions the analogy implied between the source concept MONEY and the target concept TIME holds. This often means that not all properties of the source concept are transferred to the target concept, but only a subset of them (*semantic narrowing*, Carston, 2002; Wilson & Sperber, 2004), e.g. one sense of the concept TIME is that of musical metre, which, however, is not compatible with the analogy implied between MONEY and TIME. In order to facilitate the analogy, properties may also be inferred that cannot be said to be part of either the source or the target concept, called *emergent properties* (Gineste, Indurkha & Scart, 2000): e.g. PET FISH has the property 'brightly coloured' which cannot be said to be a property of either PET or FISH. The complex meaning of the metaphor that satisfies the conditions of the implied analogy is thus *pragmatically enriched* and more than the mental contents of its constituent source and target concepts. A common way of reconciling pragmatic enrichment with the idea that concepts have core meanings is to think of the inferential process as a process of deriving an *ad hoc* concept from the source concept such that its properties match the target features of the target concept, and it is this *ad hoc* concept and not the original concept that is attributed to or included in the class, category, or domain represented by the target concept (Barsalou, 1983; McGlone, 1996; Glucksberg, McGlone & Manfredi, 1997; Glucksberg, Manfredi & McGlone, 1997). The pragmatically enriched set of inferences necessary to satisfy the metaphoric analogy, together with the respective properties of the source and target concepts (i.e. their *ad hoc* concepts), are a linguistic meaning, specifically *speaker meaning*, the context-dependent meaning intended by the speaker and the one that is most likely to be inferred by other interlocutors.

According to RT interlocutors draw those inferences during comprehension which will be helpful in understanding the speaker's intended meaning in light of the discourse context, called *speaker meaning*, as well as understanding the speaker's communicative intent in doing so. Interlocutors' assumptions about which information will be helpful is constantly updated as the discourse progresses. Linguistic meaning is thus highly dependent on the context of an utterance, the information gathered from the preceding discourse, interlocutors' conceptual and mental lexicons, and the speaker's communicative intent that the interlocutors infer. RT rejects the Gricean view that comprehending metaphors requires dedicated inferential machinery (Grice, 1975, 1978) and instead claims that there is one set of inferential mechanisms to comprehend any and all linguistic utterances. On this view, inferring a metaphoric meaning differs from other linguistic meanings in the kinds of inferences that have to be drawn in order to satisfy the implied analogy. Pilkington (2000) claims that the inferences drawn in metaphor comprehension are more likely to be a large set of weak implicatures rather than few strong inferences (e.g. entailments or strong implicatures). An implicature is that which is suggested

by an utterance but not a condition for its truth. For instance, *Paige had a baby and got back together with Walter* only implicates that her giving birth was the reason for rekindling her relationship with Walter. For all we know, Walter might not be the father of the child and so the utterance would still be true if Paige having a baby was not her reason for getting back together with Walter. According to Pilkington, the weak implicatures accumulate to produce the rich mental imagery associated with the metaphoric analogy.

According to GS the inferential process consist of two modules: (1) a linguistic module that infers salient meanings based on linguistic knowledge and (2) a pragmatic module that enriches the meaning by taking non-linguistic knowledge such as conceptual, experiential, perceptual, contextual, and world knowledge into consideration. The more familiar, common, conventional, and prototypical a meaning is, the more likely it is to come to mind more easily and readily, and the more salient it is, and the more likely it is to be inferred. Inferential strength, on this view, is then a function of *meaning salience*. According to GS, the linguistic meaning that is ultimately inferred for an utterance is one that is a compromise between the outputs of the two modules. The two modules are thought to run in parallel so that processing may align at any point during comprehension. This alignment is necessary in order to optimally arrive at a compromise. On this view, metaphor comprehension requires no dedicated cognitive mechanisms beyond the machinery of the two modules and, similar to RT, that an inferred meaning happens to be metaphoric is then a result of the inferences drawn and not the mechanisms involved. Since the second module allows context to be considered, the linguistic meaning that is inferred by the two modules working in unison is *speaker meaning*. GS's view of metaphoric meanings is thus compatible with that of RT.

CMT claims complex inferences, such as linguistic meanings, are the result of conceptual mappings from primary embodied concepts to ever more complex higher-order concepts. Which mappings are established and activated to yield complex inferences, according to CMT, depends on sensory-perceptual stimuli, in other words *context*, and the mental content and structure of concepts based on past experiences. Although CMT thinks of linguistic meanings as the result of interconnecting primary concepts to ever more complex higher-order concepts, this view is compatible with RT and GS's view of *speaker meaning* as the context-dependent linguistic meaning implied by the speaker and inferred by listeners. The complex inferences that complex higher-order concepts yield for a particular context, in CMT terms, are the complex inferences that correspond to context-dependent speaker meanings, in RT and GS terms. CMT is not just making an assumption about the cognitive mechanisms of metaphor comprehension, as RT and GS do, it is making an assumption about analytical reasoning in general because CMT takes analogical reasoning to be inherently akin to metaphor. Comprehending metaphors, on CMT's view, thus comes natural to the human mind.

Interpretations of metaphoric analogies can become conventionalised as, for instance, in the case of metaphoric proverbs such as *time is money*. A conventional linguistic meaning is also called an *idiomatic meaning* because similar to idioms, the full, pragmatically enriched meaning of the metaphoric analogy is not inferable from the linguistic form of the proverb and speakers' conceptual knowledge of the source and target concept alone. Speakers are provided with the missing information through linguistic convention and the missing information can be thought of as the context that would have to be 'spelled out' in order to allow interlocutors to infer the full, pragmatically enriched meaning in a non-idiomatic way. The idiomatic meaning of a metaphoric proverb can thus be understood as an inferential shortcut: instead of having

to consider all conceptually possible alternative interpretations and assessing their plausibility in the context of the analogy implied by the metaphor, speakers who are familiar with the idiomatic meaning are provided with a default interpretation, which eliminates the need to consider all alternatives. The idiomatic meaning therefore is a salient and preferred meaning.

Researchers in cognitive science (Lakoff & Johnson, 1980; Lakoff, 1983, 1987; Coulson, 2006) have collected substantial evidence of cross-linguistic systematicity in linguistic metaphors, which they take as evidence for (a) the metaphoric structure of the conceptual systems of the respective languages and (b) the mental representation of the respective linguistic metaphors. They claim that the structure of the conceptual system is inherently metaphoric-analogical and that therefore conceptual associations are also metaphor-like. The evidence of systematic patterns in linguistic metaphors, on this view, is taken to be indicative of the kinds of metaphoric conceptual associations that exist in the respective languages and these associations are therefore called *conceptual metaphors*. On this view, conceptual metaphors are preferred, salient meanings. They may be universal to us as a species or specific to a particular language, language family, or culture. However, many language-specific linguistic metaphors have conventional meanings called *idiomatic meanings* associated with them. On CMT's view, metaphoric idiomatic meanings are then language-specific conceptual metaphors. Importantly, this characterises idiomatic meanings as language-specific ways of conceptualising perceptual experience and introspective mental states. Contrary to CMT, RT and GS claim that idiomatic meanings are not primarily motivated by conception but by the cooperative need to facilitate communication (cf. Grice's Cooperative Principle, Grice, 1975, 1978). On CMT's view, metaphoric idiomatic meanings constitute conceptual mental representations. On RT and GS's view, metaphoric idiomatic meanings require knowledge of linguistic conventions that are distinctly *non-conceptual* because they are not primarily motivated by conceptual plausibility, world knowledge, or conceptual knowledge. With respect to metaphoric idiomatic meanings, the difference between the opposing claims made by CMT, on the one hand, and RT and GS, on the other, thus is how much influence metaphoric idiomatic meanings have on the ways that speakers conceptualise the world around them: CMT takes the influence to be rather great and largely unconscious; RT and GS take the influence to be indirectly mediated by non-conceptual linguistic conventions. We can rephrase this with respect to the evidence of linguistic metaphors: CMT takes the systematicity in linguistic metaphors to be rather direct evidence of the cognitive processes involved and of the mental structures that these processes create; RT and GS take the insight which the linguistic evidence gives into cognition to be rather indirect and mitigated by linguistic conventions.

This thesis seeks to experimentally gauge the amount of influence of non-conceptual linguistic conventions in metaphoric idiomatic meanings. In order to do this we need to set up a situation where speakers of two languages cross-linguistically have similar conceptual knowledge with regards to particular linguistic metaphors, similar expectations of what would constitute a suitable context given metaphors' conceptual properties and Fregean senses, and similar mental imagery. Among these metaphors, we then need to identify those that have idiomatic meanings which are conventionalised in only one but not the other language. Metaphoric proverbs have language-specific conventional meanings which will only be fully intelligible to speakers of the other language if they are able to infer a meaning similar to the one that native speakers have through convention (the idiomatic meaning). If CMT is correct, then non-native speakers should be able to infer meanings for the metaphoric proverbs that are similar to the idiomatic

meanings that native speakers have. If RT and GS are correct, then the meanings that non-native speakers infer for the proverbs should be different from the idiomatic meanings of native speakers and the magnitude of the difference should correspond to the size of the effect that non-conceptual linguistic conventions have on idiomatic meanings.

1.2 The experimental rationale

If inferential comprehension is primarily a conceptual process, as CMT assumes, then when speakers have similar concepts pertaining to metaphors' language-specific idiomatic meanings, they should also comprehend these metaphors similarly. On CMT's view, a metaphoric proverb's idiomatic meaning is a complex concept inferred from its source and target concept by employing a conceptual metaphor that has been conventionalised in a particular language. When two languages cross-linguistically share the embodied concepts and conceptual metaphors necessary to infer the meaning of a particular metaphoric proverb, then non-native speakers, speakers from the other language, should be able to infer the same metaphoric meaning as native speakers. In doing so, it should suffice for speakers to be led by their conceptual knowledge alone to draw plausible inferences about metaphors with L2 language-specific idiomatic meanings.

If, as RT and GS claim, comprehending metaphors with idiomatic meanings requires knowledge of language-specific non-conceptual linguistic conventions, conventions that are not motivated by or subject to conceptual plausibility, then speakers who are unaware and unfamiliar with these non-conceptual conventions should be unable to infer meanings for metaphors that are approximately similar to their L2 idiomatic meanings. In other words, if idiomatic meanings require knowledge of linguistic conventions that have nothing to do with how speakers conceptualise but which only constrain how they communicate these ideas (concepts), then in a cross-linguistic situation where we know that speakers of two language communities conceptualise in similar ways (with respect to the concepts of particular metaphoric expressions) speakers who do not know the respective L2 language-specific linguistic conventions should infer metaphoric meanings that are distinctly and significantly different from the true idiomatic meanings that speaker familiar with the conventions infer. If idiomatic meanings require knowledge of particular linguistic conventions, then when speakers do not have this knowledge, they should be unable to infer these idiomatic meanings; instead, they should infer other linguistic meanings that seem plausible to them.

The experiments in this thesis set up a situation where the chances of cross-linguistic intelligibility of metaphoric proverbs should be maximised *conceptually* so that if, as CMT proposes, cross-linguistic metaphor comprehension is primarily a conceptual process, then language-specific metaphoric proverbs should be cross-linguistically as intelligible as metaphoric proverbs which have conventional meanings in both languages. If, however, cross-linguistic metaphor comprehension is not primarily dependent on metaphoric conception but approximate similarity of non-conceptual linguistic conventions concerning their idiomatic meanings, as RT and GS propose, then cross-linguistic intelligibility should be *lower* for language-specific metaphoric proverbs than for cross-linguistically shared ones even when cross-linguistic approximate *conceptual* similarity is maximised for all metaphors in question.

Closeness in language contact and cultural interaction increases the chance of conceptual cross-linguistic similarity. The metaphors with idiomatic meanings under investigation in this

thesis are taken from English and German because these two languages are in close linguistic and cultural contact. In investigating the language specificity of idiomatic meanings of metaphoric proverbs, there are two control conditions when looking at two languages in close linguistic contact: (1) metaphoric proverbs for which speakers of both languages know idiomatic meanings and these meanings are the same for the same proverbs (*cross-linguistically shared* metaphoric proverbs), and (2) *novel* metaphoric expressions whose linguistic form resembles that of the proverbs but for which speakers of neither language have any idiomatic meanings because no such meanings have been conventionally established yet. The two control conditions are theory-neutral: Whatever the nature of idiomatic meanings is, be it language-specific conceptual metaphors or language-specific non-conceptual linguistic conventions, all proverbs categorized as ‘cross-linguistically shared’ in this thesis are classified as such because they share idiomatic meanings in the two languages. Metaphoric expressions classified as ‘novel’ have not yet been conventionally established and thus neither speakers of English nor German can have any idiomatic meanings for them.

The dependent variable, the cross-linguistic intelligibility of idiomatic metaphoric expressions, is measured in three ways in this thesis: (1) through reading/response times, (2) through metaphor plausibility judgements, and (3) through a context creation task. The independent variables are: (1) participants’ language proficiency (English monolingual native speakers, German monolingual native speakers, and English-German bilinguals), (2) the language specificity of metaphoric idiomatic meanings (English-specific, German-specific, cross-linguistically shared, and unconventionalised-novel), and (3) the mode of presentation (out of context and in context).

The rationale of measurand 1, reading/response times: If knowledge of the idiomatic meaning acts as an inferential shortcut, native speakers, who have this knowledge, but not non-native speakers because they lack this knowledge, should be faster at reading and judging metaphoric proverbs which require knowledge of idiomatic meanings. CMT would predict that because of the closeness of language contact between English and German and the exchange of concepts and conceptual metaphors that comes with it, when native speakers of either English or German make use of the same basic concepts and conceptual metaphors for a particular proverb, that proverb should be cross-linguistically intelligible to non-native speakers. In other words, non-native speakers should be able to infer a meaning similar to the one that native speakers have through convention (the idiomatic meaning). Non-native speakers should therefore not show slower reading/response times for proverbs they do not know from their native language, but read them as fast as proverbs that are cross-linguistically shared between English and German. RT and GS, on the other hand, would predict that if the idiomatic meaning is not purely a matter of conception but requires knowledge of linguistic convention, then non-native speakers who lack this knowledge should be unable to use the inferential shortcut that native speakers have through their knowledge of the idiomatic meaning and non-native speakers should therefore take longer to comprehend L2 language-specific proverbs than L1 and shared ones because they have to consider all conceptually plausible interpretations while native speakers simply default to the idiomatic meaning.

The rationale of measurand 2, metaphor plausibility judgements: Metaphors are communicative if they are intelligible and in order to be intelligible they must first be plausible (Lakoff & Johnson, 1980, 2003; Lakoff, 2008; Giora, 1997, 1998, 1999, 2003). Therefore, a metaphor whose implied analogy between source and target is implausible is also not intelligible and not communicative. The analogy is plausible only if speakers feel justified making the necessary

assumptions as part of pragmatic enrichment. Metaphor plausibility judgements can therefore be viewed as an index of their intelligibility and communicativeness. Because of the closeness of language contact between English and German and the concepts and conceptual metaphors that are shared between them because of this closeness, when speakers of both languages share the basic concepts and conceptual metaphors for a particular metaphoric proverb, that proverb should be as intelligible to non-native speakers as it is to native speakers and this, CMT would predict, should be true by virtue of conceptual plausibility. RT and GS, however, would predict that the proverb should not be cross-linguistically intelligible on the basis of conception alone, but only if speakers know the non-conceptual linguistic conventions associated with it. RT and GS would thus predict that non-native speakers should not find proverbs from the other language as plausible as proverbs from their native language because their considerations of ‘plausibility’ include their knowledge of linguistic conventions (or lack thereof) *on top of* purely conceptual considerations and they would predict that considerations of linguistic convention should take precedence over purely conceptual considerations.

Meaning is well-known to be highly context-dependent and this is recognised by all three theories, CMT, RT and GS. In this thesis, metaphor plausibility judgements are therefore measured in and out of context. Context affects the plausibility judgements of the two control conditions. Since measurand 2 defines weak and strong intelligibility relative to the two control conditions, we need to evaluate weak and strong intelligibility as the rate of change in plausibility ratings of language-specific metaphors *as a function of* the rate of change in the plausibility ratings of the control conditions, the change being the change from out-of-context to in-context metaphor presentation. CMT, RT, and GS agree when speakers should facilitate contextual information during comprehension (Lakoff & Johnson, 2003; Lakoff, 2008; Wilson & Sperber, 2004; Giora, 2003): if it yields positive cognitive effects, is informative, contributes something new to the discourse, aids comprehension, is an explanation of conceptual metaphors (i.e. it is a substitute for the conceptual metaphors that speakers might lack), and more context should make it easier to form *ad hoc* concepts. In other words, context is relevant to comprehension if it is helpful during comprehension. If contextual information is relevant for comprehending a metaphor, then speakers should make use of it. If contextual information is irrelevant for comprehending a metaphor or conceptually conflicts with what speakers already inferred, speakers should ignore this new contextual information. With enough contextual information, cross-linguistically unintelligible metaphors should become intelligible.

The rationale of measurand 3, contextual continuations: If comprehending a particular metaphor requires knowledge of its idiomatic meaning and the idiomatic meaning can be thought of as the context omitted from overt linguistic expression that would be necessary to ‘spell out’ the full conventional speaker meaning in a non-idiomatic way, then native speakers who know the intended full conventional meaning, should be able to make (at least part of) the omitted context explicit (Peleg, Giora & Fein, 2004; Peleg et al., 2008; Peleg & Giora, 2011). Since the idiomatic meaning also includes expectations as to which contextual information would be helpful during comprehension, when speakers are asked to create sensible context for a metaphoric proverb, only native speakers but not non-native speakers should be able to meet these conventional expectations. If it is appropriate to think of idiomatic meanings as language-specific conceptual metaphors, then non-native speakers should be able to make helpful context explicit on the basis of their conceptual knowledge and considerations of conceptual plausibility whenever they share the necessary basic concepts and conceptual metaphors with

native speakers. CMT would propose that idiomatic meanings are language-specific conceptual metaphors and would therefore predict that non-native speakers should be able to do this for two languages where, for the proverbs in question, basic concepts and conceptual metaphors are cross-linguistically shared because of intense language contact. RT and GS, on the other hand, see the idiomatic meaning, the knowledge of omitted context, and expectations of helpful context as part of arbitrary linguistic convention and not conception, and not necessarily as motivated by Embodied Cognition. RT and GS would therefore predict that non-native speakers should fail to make context explicit which is appropriate to the conventional meaning of an L2 proverb.

1.3 Thesis overview

A metaphor consists of two concepts, the source and the target concept, and an analogy implied between them. Chapter 2 discusses theories of concepts, in particular CMT in the framework of Embodied Cognition and its view of the inferential process as a process of dynamically deriving ever more complex higher-order concepts from primary embodied concepts through conceptual metaphors. Since I seek to investigate metaphors and their source and target concepts from a cross-linguistic perspective, the two crucial questions in this chapter are: how do we assess and compare concepts' similarity within and across languages, and how do we account for inter-speaker variation in the mental representation of concepts? Chapter 3 discusses theories of the comprehension process that infers the metaphoric analogy. I return to CMT here, and contrast it with the relevance-theoretic and graded-salience view. Of particular importance is how the nature of the inferential process changes when metaphors are not novel and unfamiliar to speakers but when they have conventionally established default interpretations (idiomatic meanings). The metaphoric proverbs in the experiments of this thesis have idiomatic meanings that are either conventionalised in both English and German or in only one of the two languages. The novel metaphors that only resemble the linguistic form of the proverbs, however, do not have any conventionally established idiomatic meanings. With respect to those proverbs whose idiomatic meanings are specific to either English or German, we can therefore test whether non-native speakers will process them as if they were novel metaphors or as if they were conventional metaphors.

It is well known that due to the history of English as a Germanic language and English's status as a global *lingua franca* in science, technology, commerce, and entertainment, English and German have one of the closest language contacts in the world. Translation of literary works, which are rich in metaphors and mental imagery, are a good measure of the exchange of cultural ideas, concepts, and mental imagery. The UNESCO's Index Translationum (2013) shows that in the last 40 years more books have been translated between English and German than between any other languages in the world. In Chapter 4, I argue that this makes English and German an ideal choice to test the predictions concerning metaphoric idiomatic meanings by CMT, on the one hand, and RT and GS, on the other, because cross-linguistic conceptual similarities should be likely between English and German. In the remainder of Chapter 4, I lay out how we may operationalise the theoretical notions pertaining to metaphoric idiomatic meanings for experimental testing.

The objective in preparing the experimental material is to maximise the chances of cross-linguistic intelligibility of metaphoric proverbs by maximising the cross-linguistic similarity of

metaphors' source and target concepts, the similarity of their contextual expectations, the cross-linguistic similarity of speakers' familiarity with the conventional form and meaning of metaphors, and the closeness of language contact. In the experiments, monolingual English native speakers will see language-specific proverbs originating from German in English translation. Conversely, monolingual German native speakers will see proverbs originating from English in German translation. We therefore need to ensure that these translations are cross-linguistically comparable in terms of the cross-linguistic similarity of metaphors' source and target concepts, the similarity of their contextual expectations, the cross-linguistic similarity of speakers' familiarity with the conventional form and meaning of metaphors. Chapter 5 focuses on the cross-linguistic comparability of the linguistic form of the metaphoric proverbs and novel metaphors used in the experiments, and Chapter 6 focuses on the cross-linguistic comparability of their linguistic meanings. Hence, Chapter 5 deals with identifying metaphoric proverbs with language-specific and cross-linguistically shared idiomatic meanings through a synchronic and diachronic corpus analysis of British English, American English, and German corpora. Chapter 6 deals with the cross-linguistic similarity of metaphors' source and target concepts and the similarity of their contextual expectations. Throughout Chapters 5 and 6 I attempt to provide both qualitative and quantitative measures of cross-linguistic metaphor similarity. When dissimilarities are unavoidable in the process of translating the material, these measures allow me to determine at the end of Chapter 6 that (a) the cross-linguistic dissimilarities are rather small and (b) they are evenly distributed across the metaphor types in the experiments (English-specific metaphoric proverbs, German specific proverbs, cross-linguistically shared proverbs, and novel metaphors).

With the chances of cross-linguistic metaphor intelligibility maximised by carefully preparing the material, the goal in the experiments is to look for lowered cross-linguistic metaphor intelligibility. Finding lowered cross-linguistic metaphor intelligibility would be evidence against CMT and evidence in favour of RT and GS. In six experiments, the degree to which L2 language-specific metaphoric proverbs are intelligible to English and German monolingual native speakers (i.e. English native speakers who do not speak German and German native speakers who do not speak English) is tested. English-German bilinguals serve as a control group: they know the idiomatic meanings of all metaphoric proverbs and should thus give baseline ratings of metaphors' cross-linguistic communicability. Since inferential comprehension is well-known to be highly context-dependent, the first three experiments (Chapter 7) test cross-linguistic metaphor intelligibility when the metaphors are presented without any context, and the other three experiments (Chapter 8) test metaphors' cross-linguistic intelligibility in context. These contexts were partially created by monolingual native speakers in the first three experiments and partially by the researcher and always with the intention to improve cross-linguistic metaphor intelligibility.

Chapter 9 summarises the three major experimental findings by considering the evidence of all three measurands together: (1) cross-linguistic communicability of L2 language-specific metaphors is lowered for monolingual speakers, (2) monolinguals and bilinguals use different heuristics in the plausibility judgement task because of bilinguals' sensitivity for metaphors' language specificity, and (3) the unexpected and counter-intuitive finding that plausibility of metaphors with familiar idiomatic meanings decreases with context. The experimental findings thus support the view of metaphoric idiomatic meanings held by RT and GS.

Chapter 2

Concepts

2.1 Introduction

A metaphor consists of (1) a source concept, (2) a target concept, and (3) an implied analogy between the two. This chapter introduces the major theories of what concepts, such as the source and target concept in metaphor, are: (1) they are mental representations of sensory perception, experience out of interaction with the world, and imagination (Section 2.2). Among these theories, of chief interest to this thesis is the view held by CMT (Section 2.3). This view combines two ideas: (1.1) all of the most basic concepts are *embodied concepts*, i.e. they are mental representations of sensory perception of either the external world as perceived through the senses or body-internal states such as emotions and intuitions, and (1.2) the mental content of a complex concept, CMT's equivalent of a linguistic meaning, is a function of its compositional structure and the contents of its constituent concepts, the smallest of which are embodied concepts. On this view, the mental content of the source and target concept plus the additional information that fleshes out the conditions under which the analogy between them is satisfied, together constitute a complex concept, the metaphor concept, the meaning of the metaphor as a whole. One of the cognitive mechanisms which CMT proposes is employed in concept composition is akin to metaphoric analogy and CMT calls this mechanism *conceptual metaphor*. CMT, thus, takes analogical reasoning to be inherently metaphoric and it takes systematic patterns in linguistic metaphoric expression to be indicative of this cognition. (2) Concepts are best understood through our ability to use them. If the successful completion of a specific task requires that we have a particular concept, then a demonstration of our ability to complete the task should be indicative of us having that concept (Section 2.4). (3) The view that concepts are Fregean sense relations, i.e. that they are abstract symbols whose meaning is defined in the propositional system of a language and by conditions of truth in the world (Section 2.5).

CMT attempts to model inferential communication, in general, and metaphor comprehension, in particular, through cognitive mechanisms that compose complex concepts from simple primary embodied concepts. On this view, *meaning salience* is the salience with which certain inferences are more likely to be drawn during comprehension. CMT uses the term *conceptual metaphors* for salient inferences. CMT views the conventionality of inferences as inherently conceptual. RT and GS, on the other hand, think of inferences as linguistic meanings where concepts are distinguished from sense relations. RT and GS therefore use a narrower definition of 'concept' than CMT. In Section (2.6), I will therefore introduce the notion of linguistic mean-

ing as used by RT and GS. This definition of linguistic meaning makes the triadic nature of linguistic symbols explicit: a linguistic symbol consists of (1) its linguistic form, (2) its linguistic meaning, and (3) the nature of the association between form and meaning (e.g. when a given linguistic form is decomposable into constituents, is the meaning of the whole inferable from the meaning of the constituents and their composition?). Metaphoric proverbs form the test condition in the experiments of this thesis. In metaphoric proverbs, the form-meaning association is conventionalised in such a way that without knowledge of the respective linguistic convention, the idiomatic meaning of the metaphor is not inferable from the meaning of its constituent source and target concepts and the analogy implied between them. On CMT's view, idiomatic meanings of metaphoric proverbs are language-specific conceptual metaphors, i.e. they are part of a speaker's conceptual knowledge. On RT and GS's view, idiomatic meanings of metaphoric proverbs are non-conceptual language-specific linguistic conventions, i.e. conventions of conceptualising that are not motivated by or may even go against pure conceptual plausibility, but are instead motivated by the need to follow linguistic convention in order to facilitate effective communication within a speech community.

The experiments in this thesis investigate the nature of idiomatic meanings of language-specific metaphoric proverbs from a cross-linguistic perspective. The rationale of these experiments is that we should choose two languages so that, for the metaphoric proverbs in the experimental test condition, the source and target concepts and mental imagery in these metaphors are *cross-linguistically similar* for the two languages. We should try to maximise this cross-linguistic conceptual similarity as much as possible. If CMT is correct, then non-native speakers should be able to infer meanings for these metaphors that are similar to the idiomatic meanings that native speakers infer. The more we maximise the cross-linguistic similarity of source and target concepts and mental imagery, the more likely, on CMT's view, cross-linguistic metaphor communicability and intelligibility should be. If, however, RT and GS are correct, then cross-linguistic metaphor communicability and intelligibility should be lowered even when metaphors' source and target concepts, including their relevance expectations and mental imagery, are cross-linguistically as similar as possible. Inter-speaker conceptual similarity is a function of speakers' individual learning path and personal experience and is thus a matter of degree. In the last part of this chapter (Section 2.7), I discuss to what extent cross-linguistic similarity of source and target concepts is attainable. If we wish to compare concepts across languages and since we rely on their representation in natural language (Fregean senses), we also need to be aware that (a) a conceptual distinction that is reflected in linguistic expression in one language, might be unmarked in another one, and vice versa, (b) an apparent difference in linguistic expression between two languages might not actually indicate a conceptual difference. We should choose metaphors for the experiments of this thesis where, for English and German, their linguistic form relates unambiguously to its idiomatic meaning.

2.2 Concepts as mental representations

The view that concepts are mental representations of the things we *perceive* and *imagine* has a long history. One major development since antiquity has revolved around the distinction between those mental representations that refer to external objects, accessible to us through sensory perception, and those mental representations that are the product of our imagination. The Stoics (301 BC to the 2nd century AD) did not distinguish sensory perception and imag-

ination yet, but they distinguished *apperception* from *perception* as those perceptions that are confirmed to be real through analytic proof. René Descartes (1641), on the other hand, used the term *apperception* as perception through imagination and the senses. Gottfried Wilhelm Leibniz (1686, 1704, 1714) was among the first to recognise that if parts of our perception are unconscious, parts of our mental representations might be unconscious, too. He distinguished between *perception* (sensory perception), *small perception* (un/subconscious sensory perception), and *apperception* (the conscious mental process of analysing what is consciously perceived).

With the advent of the psychology of perception and modern neuroscience, there has been a revival of the view that concepts are rooted in sensory perception and there are attempts to ground a theory of the conceptual system on our understanding of the capabilities of its neural basis. In effect, the mental representations view has become the *de facto* standard view in cognitive science and psychology, and it has widespread support in linguistics and the philosophy of mind and language (Barsalou, 1999; Millikan, 2000; Carruthers, 2000, 2006; Margolis & Laurence, 1999, 2010; Fodor, 1983, 2008). What connects the modern strands of the mental representations view is their commitment to empiricism. I will not discuss approaches that reject empiricism and study concepts through *a priori* introspection in this thesis. In this section, I discuss some of the major strands and critiques of the representational view before I focus on CMT in the next section. Since its beginnings in the 1980's (Lakoff & Johnson, 1980; Lakoff, 1983, 1987), where CMT studied the conceptual system primarily through conceptual analysis, it has changed to a neural theory (Lakoff, 2008, 2009), which studies the conceptual system from a cognitive perspective. Because of this theoretical shift, I will point out some relevant insights from neuroscience and the psychology of perception, so that we can evaluate the claims of the neural theory of Conceptual Metaphor on this basis.

A common definition of concepts in theories that see them as mental representations describes them as *mental images*. John Locke (1690) and David Hume (1739), among others, for instance, used the term *ideas* or *mental images* for concepts. The claim can certainly not be that *concepts* are always visual mental images or that *all* mental representations are mental images. We have to understand *image* in a weak sense here: 'image' as in *imagination*. Imagination can come in many modes, not just visual. Painters might imagine visually, sculptures volumetrically, musicians acoustically, athletes kinetic-motory. In fact, neurophysiological evidence shows that when people of these respective groups are asked to imagine their creative process, imagination runs on the same neural circuitry that would be active if they were actually performing that activity. The same brain regions are activated during sensory perception and imaginative recall of visual imagery (Ganis, Thompson & Kosslyn, 2004; Gelbard-Sagiv, Mukamel, Harel, Malach & Fried, 2008), graspable objects (Creem-Regehr & Lee, 2005), actions and movements (Decety, 1996; Gerardin, Sirigu, Lehericy, Poline, Gaymard, Marsault, Agid & Le Bihan, 2000), acoustic sounds (Zatorre & Halpern, 2005; Bunzeck, Wuestenberg, Lutz, Heinze & Jancke, 2005). Empirical evidence thus shows that imagination runs on the same neural circuitry as actual sensory experience and action. In addition, evolutionarily, the neural circuitry of the conceptual system was designed for visual perception; the term *mental image* could thus also be understood in this sense. If we can construe a definition of concepts as mental representations, the next question must concern the relationship between concepts, i.e. the structure of the conceptual system.

One if not the central idea that has emerged since the 1950's is that in the conceptual

system concepts do not just relate to one another as atomic units, but they are intertwined into ever more complex concepts. The way that these higher-order structures emerge from basic concepts varies between theories: in Jerry Fodor's (1975; 1987; 1990; 1991; 1994; 2008) *language of thought*, for instance, it is propositional, in Noam Chomsky's (1996) *i-language* it is similar to syntax in natural language, while Lakoff (2008, 2009) sees it as a neural network. As we will see shortly in our discussion, 'propositional', 'syntax', and 'network' are not competing views about the structure of the conceptual system, they are different modes of describing its structure. Those theories that are in pursuit of a neurophysiological grounding of their claims, such as the neural theory of Conceptual Metaphor, see the sub-symbolic level of concepts as patterns of activity in the neural network. The central tenet, in which I use the term 'compositional' to describe the nature of the structure, is the tenet that all mental representation views share:

The content of a complex concept is a function of its compositional structure and the contents of its constituents.

Ludwig Wittgenstein (1953, 1958) had an important insight into the structure of the conceptual system: *family resemblance*. He gives the example of the concept GAME. How do we define this concept, if, as Wittgenstein assumes, they have no common denominator? Some games are competitive, others cooperative. Some can be played alone while others can only be played together, and so on. The important point about Wittgenstein's insight is that when the central tenet says that a complex concept is a function of its compositional structure and the contents of its constituents, we cannot take it to mean that this is an *inheritance network* where the complex concept inherits all features of its constituents or where, conversely, we can assume that there is one feature of the complex concept that *all* of its constituents also have. Wittgenstein's family resemblance shows that one can envision a network of interconnected concepts where a concept that is a constituent in the complex concept shares only some features with another constituent concept which again shares some *other* features with the next constituent concept and so on, until all constituent concepts form a chain in this way, where none of them have a common denominator, yet they are grouped through interconnection. This chain structure is one way of many to connect them and to think of them as a complex concept. Hence, Wittgenstein's contribution is to show how important it is to consider the *structure* of the network of constituent concepts, the structure of the interconnection that is constitutive of the function that *is* the complex concept.

Eleanor Rosch (Rosch, 1973, 1975; Rosch, Mervis, Gray, Johnson & Boyes-Bream, 1976; Rosch & Mervis, 1975; Rosch, 1983, 1978) contributed the important insight that it can be shown experimentally that basic level concepts invoke mental images, activate motor programmes, and allow Gestalt perception, while more abstract concepts do not: For instance, we can have a mental image of a prototypical chair or a typical table or a generic cupboard, but we cannot conjure up such an image for *furniture*. We can also pick these objects out through Gestalt perception at a distance. And we associate a motor programme with each of them, a set of actions that when carried out involve the objects, e.g. you sit in a chair, you put food on the table, and you can take things out of the cupboard. Similarly, the words *car*, *bicycle*, *boat*, *plane*, and *space craft* might each invoke a mental image, yet we cannot conjure up an image of a generic *vehicle* that combines all the features of cars, bicycles, boats, ships, planes, and space crafts. We can tell a car from a boat or a plane through Gestalt perception, even if within each type their design may vary greatly, but we have no Gestalt perception for the more

abstract concept of VEHICLE. And with each of the basic level concepts we associate motor programmes, e.g. you drive a car, driving a bicycle involves pedalling, you steer a boat, and you fly a plane, but there is no motor programme for the abstract concept VEHICLE. The same applies to the abstract concept of MACHINE that subsumes all types of mechanical and electronic machines. The important insight here is that concepts are not autonomous units in the conceptual system, they are interconnected and, as in the case of Rosch's examples of furniture, vehicle, and machine, multiple connections can be activated simultaneously and, at least, for some concepts it is this simultaneous activation that is constitutive of them being concepts in the first place. These concepts are a function of that simultaneous activation. The examples also show that for basic level concepts the brain must have associations between mental imagery, Gestalt perception, and motor programmes, i.e. their nature is defined by that association.

Exceptions to the central tenet are, for instance, cases where complex concepts have features that are not part of the contents of their constituents. Because these features do not come from the constituents, but emerge through the process that creates the complex concepts, they are called *emergent properties* (Gineste et al., 2000). For instance, Margolis & Laurence (2011) give the example of a PET FISH which has the feature 'brightly coloured' associated with it, but this feature cannot be said to be necessarily part of either the concept PET or the concept FISH. It must therefore be an emergent property. We can account for emergent properties in the definition of the central tenet if we can find a way how emergent properties are created as part of the function that gives rise to the complex concept. As I am not interested in developing such an account in this thesis, I shall omit its development. In all of the examples, experimental material, and results I will comment on potential emergent properties whenever they should arise and I will comment on how they arise by showing how the interconnection of concepts and the derivation of complex concepts might motivate such emergent properties. Similarly, Fodor & Lepore (1996) use the examples PET FISH and RED HERRING to argue against prototype theory.

Prototype theory (Osherson & Smith, 1981; Jones, 1982; Smith & Osherson, 1984; Osherson & Smith, 1988; Storms, De Boeck & Ruts, 2000) defines concepts as categories. Inclusion in categories is determined by the degree to which a particular object matches the features of a prototypical member of that category. The overwhelming empirical support makes prototype theory the *de facto* view in cognitive science and psychology. Critics of prototype theory (Margolis, 1994; Fodor & Lepore, 1996) commonly point to atypical examples of a category: Is the pope a bachelor? Is a dog that lost its leg still a dog? What if it has been surgically altered to look like a raccoon? Corals are animals, not plants, even though they are stationary. Plants can move voluntarily which we usually associate only with animals. For instance, plants turn their leaves to allow for optimal exposure to sunlight, they close their blossoms over night and open them again in the morning, and the parasitic dodder *Cuscuta*, also known as strangleweed, lovevine, and hellbind, does not put down roots and is not capable of photosynthesis. It relies on its host plant for nutrients. It does not attach itself to any host plant in its proximity but only to hosts which are healthy, a behaviour more akin to predators in the animal kingdom. World knowledge and how specific we want to be in a particular situation determine our categories and prototypes: for instance, if we know that the genetically reconstructed earliest ancestor of all life on Earth was an organism capable of photosynthesis, which we usually associate with plants, *and* with a cell membrane we typically associate with animals, then corals might not

seem so exotic an animal after all.

We might take Rosch (1975, 1983) to provide counterexamples to prototype theory: we can think of a prototypical table or chair or cupboard, but we cannot envision a prototypical piece of furniture that would combine and blend all of their features. Similarly we might be able to have a prototype of a car, which abstracts away from atypical examples such as stretch-limousines and pick-up trucks, or we might have a prototype of a plane, which abstracts away from the fact that not all planes are motorised or some planes are single-winged, or we can have a prototype of boats, which abstract away from the fact that boats do not necessarily have sails, are motorised, or require rudders or oars. However, we cannot have a prototype of a vehicle that combines all of the aspects of cars, planes, and boats. Even though our world knowledge might tell us that there are, for instance, amphibian vehicles that can move both on land and water or attempts of cars that can fly, most people would intuitively categorise these as atypical examples, in the same way that a surgically altered dog is atypical, and would not think of them as prototypical vehicles. We might reply to this by saying that prototypes are learned culturally and so if as a culture as a whole we agree on certain prototypes, then when language learners encounter things such as amphibian vehicles they try to fit them into the categories they already have, which in effect makes them atypical because they will not fit, neither in the category of cars nor the category of boats.

We might elaborate on this critique of prototype theory on cultural grounds and say that, for instance, people in the past might have had very different prototypes of automobiles, planes, or boats. Think of early bicycles where the front wheel is much bigger than the back wheel or the prototypical boat that Ancient Egyptians, Vikings, 17th century explorers, or we would envision today. Or how before the invention of the printing press the concept of WRITING might have invoked mental images of quills and ink, while after the invention of the printing press we might also include typing in our concept of writing, and in the case of a tablet computer, writing might not even involve a pen or ink or mechanical keys. Hence our modern-day concept of writing would be very different from that of people in the past. Similarly we might say we see the same variation of concepts geographically that we see historically: Someone in Europe might have a very different prototypical idea of an ANIMAL than someone living in Australia because the representatives of the concept ANIMAL, their perception, would lead to different mental representations. The important insight, however, is not that categories are tailored and vary with world knowledge and communicative intent; the insight is that no matter how we divide the categories, we think in categories, features, typical and atypical features, and category members in the first place—at least for some concepts.

Another critique of the representational view of concepts concerns the question of whether emotions are concepts or not (Wierzbicka, 1992). CMT takes many of the most basic concepts to be rooted not just in perception, but in emotions. Zoltán Kövecses (1990, 2000, 2002) takes a similar position. António Damásio (1994) shows that patients who as a result of brain injuries and strokes can no longer experience emotion also are less able to set goals for themselves and make plans because they can no longer reason about what would be good for themselves and what would be good for others. Many 17th century thinkers such as Descartes, Hume, and Leibniz, along with the Stoics, thought that emotion impedes reason. This modern physiological evidence, however, suggests that emotion is necessary for reason. Lakoff gives the example of the metaphor AFFECTION IS WARMTH and argues that this is a primary metaphor concept which results from emotional experiences in early childhood. Hence, CMT, as a particular

strand of the mental representations view, has already shifted its definition of *concept* from a classical to an embodied-cognitive notion which in its current form already incorporates emotion concepts because it takes emotion to be necessary for reason.

Some (Heck, 2000; Kelly, 2001; Gunther, 2003; Lerman, 2010) have proposed that not all of cognition might involve concepts, but that there can be non-conceptual cognition. This might, for instance, be the case where an ability is acquired solely through imitation or where knowledge is not explicit. Although I am generally open to this possibility, I shall not discuss it further because this thesis focusses on the nature of complex (metaphoric) concepts and their relation to one another. Non-conceptual cognition, if it exists, is therefore not of interest here. Notice, however, that non-conceptual cognition might nevertheless be rooted in sensory perception, e.g. learning solely through internalised imitation, thereby suggesting that they would also be mental representations, albeit non-conceptual mental representations.

By far the biggest problem for the mental representations view are concepts without reference (or possibly even without perception): fictional names of mythical creatures such as *hobbit*, *dragon*, *unicorn*, *cyclops*, or *phoenix* have no referents in the real world which could be perceived to create the mental representations necessary for their concepts. If *all* concepts are rooted in perception and there is no external referent to be perceived and pointed out as *a hobbit*, *a dragon*, etc., then how are we to acquire these concepts? In this case we might have to retreat to saying that they are not concepts. I would argue that there *are* external referents to be perceived. Of course these creatures do not really exist, but there are cultural ideas and representations of them, depictions in art and literature, and those are the representations that form the basis on which we acquire these concepts. Fregean sense relations, which I will discuss in more detail in Section (2.5), offer an alternative account. On this view, we can say that there is a concept HOBBIT, which as its extension may not map to a creature in the actual world, but which nevertheless has a word, a linguistic symbol, which can be used to refer to the concept. Hence, the word *hobbit* has a sense and as a linguistic symbol stands in relation to the rest of the grammar: for instance, we know it is a noun in the grammar of English, it refers to a living agent, and so on, and this grammatical knowledge is part of the *sense* of the word that relates it to the concept HOBBIT. The problem that these examples pose is that they call for a general treatment of *fictionality* in our theory of concepts. Large parts of human language, communication, and artistic linguistic expression are not concerned with the sort of factual truth that is at the heart of truth-conditional treatments of concepts. As a result they pose a problem for any propositional treatment in formal semantics and strands of the mental representations view that think of the structure of the conceptual system as being propositional. Lewis (1978) insightfully points out that there are multiple degrees of fictionality: The fictional world of Sherlock Holmes, for instance, despite its fictional deviations from the real world, is much more similar to the actual world than that of, say, the creatures of Greek mythology. The mythical creatures have abilities that we know physically to be impossible while the world of Sherlock Holmes adheres to the physics of the real world. Possible Worlds Semantics offers a comprehensive treatment of fictionality. It traces back to the work of Carnap (1947) and was fully worked out by Hintikka (1957, 1961), Bayart (1958, 1959), Kripke (1959, 1963a,b, 1972), and Lewis (1986). I will return to non-referential concepts later on in this chapter (in Section 2.6), however, since there are no examples of this kind in the experiments that I present in this thesis, we shall not concern ourselves with resolving this issue.

In the metaphoric proverbs used in the experiments of this thesis we find no fictional con-

cepts. The only two emotion concepts in the experiments are LOVE and JEALOUSY. We might consider emotion concepts to be problem cases: what are the features of emotions if emotions are also concepts? If emotions are qualitatively different from other concepts, what happens when the implied metaphoric analogy is between an emotion concept and a non-emotion concept, e.g. AFFECTION IS WARMTH, LOVE IS BLIND? In AFFECTION IS WARMTH, AFFECTION is an emotion and WARMTH is a physical property. In LOVE IS BLIND, LOVE is an emotion concept and BLIND is the property of the concept BLINDNESS. I would argue that even if emotion and non-emotion concepts are qualitatively different types of concepts, empirical evidence of proverbs, such as the ones under investigation in this thesis, suggests that attribution/analogy implication is possible and conceptually successful because the metaphoric meanings of these proverbs are conventionalised as idiomatic meanings as part of the same conceptual system that the constituent source and target would be part of. At least in the scope of this thesis I therefore consider emotions not to be problematic for a theory of concepts. In the next section, we look at Conceptual Metaphor Theory (CMT), which takes concepts to be mental representations, the most fundamental of which are rooted in perception and experience of body-internal introspective mental states. The view that thought cannot be properly understood without understanding its roots in bodily experience is commonly called Embodied Cognition. The embodied-cognitive view is by no means exclusive to CMT. Its proponents come from many different disciplines, among them linguistics, philosophy, psychology, and cognitive science. As diverse as the host disciplines are the approaches to Embodied Cognition. Given this diversity, we forgo a discussion of all of them and concentrate on the embodied-cognitive positions of CMT (Lakoff & Johnson, 1980, 2003) and Raymond Gibbs (2003, 2006) which will be of interest to us in this thesis.

2.3 Embodied Cognition and Conceptual Metaphor Theory

2.3.1 Inferential comprehension as creating ever more complex higher-order concepts from primary embodied concepts

CMT is phrased in neural terminology to emphasise that the object of study is, as George Lakoff (2014) calls it, *real reason*, the way that the human mind really reasons within the biological constraints of the brain, rather than some abstract, idealised, logical theory of reason which hypothesises without empirical physiological constraints in mind. A basic understanding of the nervous system is therefore necessary.

Nerve cells are called *neurons*. The connections between neurons are called *axons*. The point where an axon from one neuron meets another neuron is called a *synapse*. Synapses are the place where the brain “makes decisions” because here they decide when and whether to transmit information. Information is sent through the neural network in the form of electric signals. Axons are not in direct physical contact with neighbouring neurons, which, however, would be needed to transmit electric signals. A synapse is thus a gap between an axon and a neighbouring cell. To send information across this gap, synapses use the transport system that underlies cellular metabolism. In the cellular transport system, nutrients are released automatically upon arriving at their destination, but in the synaptic transport system there is

a trigger mechanism that decides *when* to release neurotransmitters, the chemicals inducing an electric signal in the neighbouring neuron, thereby transmitting information. For completing our understanding of the trigger mechanism, Thomas Südhof was a co-recipient of the 2013 Nobel Prize in Physiology and Medicine. We are born with a fixed number of neurons. Cells that die as a result of traumatic injury are rarely replaced by new cells, but when repair occurs, replacement does not happen quickly enough to allow recovery of the lost circuitry, rather the new circuitry is a bypass (Hampton, Innes, Merkler, Zhao, Franklin & Chandran, 2012). Learning, such as concept acquisition, is the growth of axons. In the same way that synapses “decide” when and where to transmit information to, they “decide” when to “feed” connected neurons. The neural circuitry thus controls the growth of its own interconnections (Cao, Maximov & Südhof, 2011).

Central to the idea of real reason is that the biology of the brain limits the capabilities of human cognition. There is a finite number of neurons and a finite number of axons. The average baby has about 100 billion neurons at birth. In the womb and after birth up to around one quadrillion axons are established. Synapses can only fire every 1 to 5 milliseconds because that is how long it takes to reset the trigger mechanism. Therefore there is a finite limit to the amount of information the brain can process simultaneously and so there is a limit on how many thoughts or complex a thought we can entertain at any one time. The brain regulates all bodily functions and so the portion of the network that is available for higher cognitive functions, such as the conceptual and inferential processing we are interested in, is limited and decreases with physical stress. Lakoff (2014) points out that the temporal threshold where sequentiality becomes conscious is about 40 milliseconds. Synchronised connected neural activity that occurs within this 40 milliseconds window is perceived as one inference. The 40 milliseconds threshold fits well with psycholinguistic evidence, e.g. the P200 and N400 spikes in EEG or reaction times of linguistic intuitions such as grammaticality and acceptability judgements. The threshold on sequentiality is also a limit of introspection because if we perceive all inferences occurring within 40 milliseconds as one, we will be unable to analyse them introspectively.

Sensory perception and perception of body-internal states is sent to the brain where it is represented in *topographic maps*. The idea of topographic maps goes back to Martha Farah (2004). Through the work of Michael Silver and his lab (Silver & Kastner, 2009) we know, for instance, that if one shines a light across the retina, the neurons in the corresponding topographic map are activated one after the other, thereby retracing the movement of the light source. Arrays of neurons in the sensory system are mapped onto arrays of neurons in the brain, thereby creating a neural map of what the sensory system perceives. On CMT’S view, topographic maps then map to *image schemas*. Image schemas were discovered by Ronald Langacker (1982, 1987a,b, 2008) and Leonard Talmy (1983). They allow to filter and abstract away from the perceptual representation of topographic maps. For instance, schemas identify semantic roles such as INTERIOR, EXTERIOR, CONTAINER, PATH and GOAL by creating connections with the corresponding regions in a topographic map and suppressing activity in other regions.

More complex concepts, Lakoff (2009, 2013) claims, are combinations of many topographic maps and image schemas in what he calls *neural cascades* (X-nets, the term that Srinivas Narayanan used in his dissertation for interconnected layers of topographic maps, Narayanan, 1997). The conceptual system is then seen as a network of interconnected cascades and calling some portion of this network, some substructure of it, one concept or another becomes an arbitrary

secondary question of convention on CMT's view. This characterises CMT as a theory of Embodied Cognition because it takes all concepts to be ultimately rooted in sensory perception. Inferences often involve abstract concepts. In light of CMT's strong claims about embodied mental representation of concepts, it has an interesting approach to *abstraction*. CMT argues that the more primary concepts are activated simultaneously, the more abstract the complex composite concept intuitively feels. Therefore, CMT argues that even the most abstract concepts are ultimately grounded on embodied perceptual primary concepts. On CMT's view, meaning is thus always embodied and embodiment is the way that CMT accounts for the contextuality of the inferential process. CMT argues that what we perceive in a given context stimulates a particular primary firing pattern in topographic maps which then inhibit all of the parts in complex cascades that do not fit with that context, i.e. what prototype theory would call *atypicality*, GS refers to as *saliency*, and RT as *contextual relevance* (see Chapter 3). This means that inferences, on CMT's view, are *speaker meaning*, the meaning implied by a speaker's utterance and inferred by listeners in light of its discourse context. CMT, RT, and GS thus agree on the notion of speaker meaning.

In neuroscience, it is hypothesised that concepts and their mental models might correlate to *cell assemblies* of neurons, a term introduced by Donald Hebb (1949). On CMT's view, acquiring a concept means to fire and associate a certain set of synapses with it upon exposure to a contextual-perceptual stimulus. Accessing a concept means firing the corresponding neural circuitry. When multiple concepts need to be accessed simultaneously or in quick succession, the simultaneous or sequential firing needs to be coordinated. The different concepts needed might be located in different places of the network, so in order to coordinate their firing, neural circuitry between them needs to be recruited to establish a temporary information transfer, a process called *neural recruitment*. When the synchronised firing of the circuitry of multiple concepts requires less than 40 milliseconds, they are perceived as one concept. For instance, Terry Regier (Regier, 1991; Regier & Carlson, 2001) and Lakoff (2009) argue the image schema of the concept INTO can be created by mapping the image schemas of IN and TO onto each other. The image schema of IN has the notion of a container which in turn has an interior and an exterior. The image schema of TO is about motion. It has a source, a path, and a goal. In order to get the image schema for INTO from the image schemas of IN and TO, the elements of those schemas have to be mapped onto each other: The exterior of the container is the source, the interior of the container is the goal, and the motion from the exterior to the interior is the path. The connection between the schema of IN and TO cannot be permanent, otherwise we would not get IN and TO individually, but always activated in unison as INTO. Synchrony in neural recruitment is thus a prerequisite for temporal-sequential compositional inferencing.

Around the age of five half of the axons, those least used, die off. CMT calls the concepts acquired up until this dying-off *primary concepts*. Lakoff (2008, 2009) gives the example of the primary metaphor AFFECTION IS WARMTH. He argues that since the brain is always computing temperature but not always computing affection, more neural activation will flow from the temperature region in the brain to the affection region than the other way around. This, he argues, is why AFFECTION IS WARMTH seems intuitively more plausible than WARMTH IS AFFECTION. Past the age of five *all* learning is achieved through neural recruitment because by then the locations of primary concepts in the network have been fixed. An inference involves accessing multiple concepts, whose coordinated activation is achieved through neural recruitment. CMT thinks of *meaning saliency* as the strengthening of frequently recruited neural

connections. Strengthening causes temporarily recruited circuitry to become permanent, a process called *neural binding*. Conventionality of inferences is thus seen as inherently conceptual. There are many types of interconnecting multiple concepts. The type used in metaphor comprehension matches metaphors' source concepts to their corresponding target concepts through a process called *similarity matching*. Similarity matching is matching the features of the source concept to those features of a target concept that are most similar to it.

There is substantial behavioural evidence (Kahneman, 2002; Tversky, 2004) that in categorisation and attribution tasks participants employ similarity matching which suggests that it is fundamental in analogical reasoning. CMT also refers to similarity matching as *conceptual metaphor* because of the striking similarity to metaphoric analogy. CMT attempts to model inferential communication, in general, and metaphor comprehension, in particular, through cognitive mechanisms that compose complex concepts from simple primary embodied concepts. In metaphor comprehension, CMT argues, conceptual metaphor is the dominant process guiding the inferential process. Conceptual metaphors are permanent neural bindings of the neural circuitry corresponding to metaphors' source and target concepts. According to CMT, conceptual metaphors are therefore the driving factor of *meaning salience* in metaphor comprehension. Meaning salience is the salience with which certain inferences are more likely to be drawn during comprehension. CMT thus takes analogical reasoning to be inherently metaphoric and it takes systematic patterns in linguistic metaphoric expression to be indicative of this cognition. In CMT, the terms *metaphor circuitry* and *conceptual mapping* are also used interchangeably with the term conceptual metaphor.

Lakoff (2009, 2014) characterises *metaphor mapping* as synchronised coordinated firing that maps from one schema, in one part of the brain, to another schema, in another part of the brain, with corresponding semantic roles. The activation of whole collections of neurons is coordinated by Gestalt circuits. Gestalt circuits are collections of neurons that are activated together, i.e. one neuron activates a whole set of other neurons, for instance, to trace the motion path of a light source. Most *complex metaphors*, Lakoff argues, are combinations of primary metaphors. They are combined through neural binding of multiple primary metaphors. So cascades of simple metaphors can be recombined into higher-order cascades that correspond to complex metaphor concepts. According to CMT, inferring metaphoric idiomatic meanings is primarily a matter of conception. According to CMT, conceptual metaphors can be (1) language-specific, (2) shared among several languages (e.g. within a language family with a shared cultural history), or (3) they can be universal (i.e. occur species-wide around the globe). The idiomatic meanings of metaphoric proverbs are language-specific conceptual metaphors.

Sam Glucksberg (Glucksberg & Keysar, 1993), Matthew McGlone (2007), and Boaz Keysar have long argued and provided experimental evidence against CMT. Keysar, Shen, Glucksberg & Horton (2000) argue against the existence of conceptual mappings such as the ones proposed by CMT on the basis of their experiments. Indeed, in much of their experiments they argue against particular kinds of conceptual metaphors proposed. For instance, Glucksberg, McGlone & Keysar (1992) found that when participants were asked to interpret the three metaphoric linguistic expressions *Our love is a bumpy roller coaster ride*, *Our love is a voyage to the bottom of the sea*, and *Our love is a dusty road travelled*, which Gibbs (1992) takes to be indicative of the conceptual metaphor LOVE IS A JOURNEY, they arrived at very different interpretations which often made no reference to the supposed conceptual metaphor. Another example, where instead reading time is used as a measurand, which should be more intuitive than the reflective

interpretations in Glucksberg et al. (1992), is Keysar et al. (2000). They assume that when particular conceptual mappings are required in order to comprehend a linguistic expression, then when these mappings are not available to speakers, they should take longer to comprehend, and they show experimentally that participants' reading times, however, do not vary significantly when reading the two kinds of expressions. They therefore conclude that conceptual metaphors of the kind that CMT proposes might not exist.

2.3.2 Metaphor systematicity as evidence for the structure of the conceptual system and of meaning salience

Systematic patterns in linguistic metaphors emerge within and across languages. The systematicity within languages is staggering by any measure and the systematicity in which ways of conceptualising seem to appear in particular languages but not others have motivated some researchers (Lakoff & Johnson, 1980; Lakoff, 1983, 1987) to claim that these patterns are indicative of differences in the conceptual systems of speakers of the respective languages. CMT takes these systematic patterns of metaphoric expression to be indicative of cognition. Systematic patterns in linguistic metaphor are well known to exist in all languages (Coulson, 2006). Consider the following examples from English:

- (1a) Her thesis rests on a firm foundation.
- (1b) His argument crumbled and broke down.
- (1c) I rest my case.
- (1d) A conclusion is built on premises.

These examples seem to follow an underlying logic: ARGUMENTS ARE BUILDINGS, i.e. that arguments are like buildings. We also find that there are often different metaphors that allude to the same abstract and complex concept. Here are some examples that Lakoff notes for the complex concept of THOUGHT in English:

- (2a) THINKING IS MOVING
- (2b) IDEAS ARE LOCATIONS
- (2c) COMMUNICATION IS LEADING
- (2d) UNDERSTANDING IS FOLLOWING

- (3a) THINKING IS PERCEIVING
- (3b) IDEAS ARE THINGS SEEN
- (3c) COMMUNICATION IS SHOWING
- (3d) UNDERSTANDING IS SEEING OR PERCEIVING

- (4a) THINKING IS OBJECT MANIPULATION
- (4b) IDEAS ARE OBJECTS
- (4c) COMMUNICATION IS SENDING
- (4d) UNDERSTANDING IS GRASPING THESE OBJECTS

- (5a) THINKING IS EATING
- (5b) IDEAS ARE FOOD
- (5c) COMMUNICATION IS FEEDING
- (5d) UNDERSTANDING IS DIGESTING

Moreover, metaphors come in hierarchies from concrete to abstract with ABSTRACT IS CONCRETE as the root, i.e. abstract concepts such as ARGUMENT can be understood metaphorically, by means of analogy, in terms of concrete concepts such as BUILDING. Multiple alternative metaphors that allude to the same complex concept also fall somewhere along this abstract-to-concrete hierarchy. In the case of the example of THOUGHT we see that all of the English metaphors follow the ABSTRACT IS CONCRETE pattern: THINKING is abstract, MOVING is concrete; IDEAS are abstract, LOCATIONS are concrete; COMMUNICATION is an abstract concept, LEADING is a concrete concept; UNDERSTANDING is abstract, and FOLLOWING is concrete, and so on. So similar to hyponymic taxonomies (e.g. a semi-detached is a type of house and a house is a type of building and a building is something built by humans, rather than by nature, and so on) metaphors also come in hierarchies according to the abstractness or concreteness of the conceptual mapping. And we need to keep in mind that this cross-linguistic evidence of the ABSTRACT IS CONCRETE pattern is taken as evidence in CMT that *all* abstract concepts are understood metaphorically, i.e. computed in the brain, using embodied-cognitive primary concepts.

The question concerning the empirical evidence of this systematicity is whether it is indicative of cognition. Does the fact that we systematically structure metaphoric utterances mean that the patterns that emerge from that systematicity are present in speakers' minds prior to linguistic expression? The Sapir (1921)-Whorf (1956) hypothesis states that an individual's thoughts are shaped by the language(s) that individual speaks natively. Whorf (1940) gives an example of a cross-linguistic difference in gestalt construction. An expert gunman would know that firing a gun leaves burned gunpowder residue in the barrel, which, if not removed properly, will degrade the gun's precision. The expert gunman would also know that in cleaning the barrel, one must avoid leaving scratch marks because scratches on the inside of the barrel would also degrade the gun's precision. Since most guns have long barrels, the brush used for cleaning is often mounted on a rod. Removing the residue burned to the metal can be difficult and so the rod needs to be forced into the barrel, which is why it is also called a ramrod. If we wanted to instruct a gunman to remove the burned gunpowder residue from the inside of the gun's barrel, English speakers might say *Clean it with the ramrod*. Sapir (1921) contrasts this with the Shawnee sentence *Nipēkwālakha*. There is only one proper way of cleaning the barrel, but the English and Shawnee utterance employ different gestalt constructions to invoke it. The English construction uses the gestalt concept of CLEANING, the preposition *with* which indicates that the RAMROD is the instrument used for cleaning. The Shawnee construction uses the gestalt concept of a DRY SPACE (*pēkw*), the INTERIOR OF A HOLE (*ālak*), and the manner of cleaning by MOTION OF THE TOOL or INSTRUMENT (*h*). English and Shawnee gunmen would clean the gun in the same way which demonstrates that they conceptually agree on the action prompted by the utterances.

Speakers' knowledge about guns includes that they have barrels. The English gunman would know that a ramrod, essentially a brush on a rod, is only used for cleaning the inside of the barrel. So for the expert gunman, the gestalt concept RAMROD is enough to indicate that the object

to be cleaned is a gun and that the part of the gun to be cleaned is the inside of the barrel. Mentioning the gestalt concept of a RAMROD, GUN, and the INSIDE OF THE BARREL would be redundant. A gun's barrel is essentially a hollow tube. So the Shawnee gestalt concept INSIDE OF A HOLE is actually quite an accurate characterisation of a gun's barrel. The expert gunman would know that the barrel should only be cleaned with a ramrod. So although the Shawnee utterance does not mention the RAMROD, the Shawnee gunman would know this is the tool to be used for cleaning. So the English utterance uses the RAMROD gestalt concept to specify the object, place, and manner of cleaning; the Shawnee utterance indicates the tool to be used by specifying the place and manner of cleaning. Notice that both the English and the Shawnee gestalt constructions underspecify, i.e. omit information, because when the addressees of the utterances are expert gunmen, the speaker can assume that they will know the proper way of cleaning the barrel from burned gunpowder residue. We routinely underspecify in our use of natural language and rely on interlocutors conceptual knowledge to fill in the gaps. Imagine how detailed the cleaning instructions would have to be if we knew the addressee to never have handled a gun. When utterances focus only on a subset of the conceptual features in gestalt construction, then there are always multiple possible gestalt constructions for the same conceptual knowledge. Alternative gestalt constructions for the same conceptual knowledge do not only occur in the cross-linguistic case. Notice that English speakers might prompt the gunman to clean the gun by saying *Clean the barrel of the gun*, where the ramrod is not mentioned, but, similar to the Shawnee utterance, the part of the gun to be cleaned is mentioned, which in turn implies that a ramrod should be used. For an expert, the English sentence *Clean the gun*, which does not mention the ramrod, is informative enough for just the same reason that the Shawnee sentence is. One might therefore wonder why Whorf (1940) insists that English and Shawnee speakers use different gestalt concepts. The sentence *Clean the gun* should invoke similar gestalt concepts for expert English and Shawnee gunman. Whorf's argument is not that gestalt conception did not overlap cross-linguistically, he argues that cross-linguistic differences in gestalt conception can occur, and we can grant him that much.

Kay & Kempton (1984) argue the extreme version of the hypothesis, that *all* thought is constrained by language, has been disproved. We also see from the gun cleaning example that the hypothesis cannot be that language shapes thought in such a way that there are certain thoughts that speakers of one language can entertain which, however, cannot be understood by those who live in another language. The opposite extreme, that language does not influence thought at all, is also widely considered to be false because certain gestalt constructions make particular features of a concept more salient than others. In order to be economic in our linguistic expression we routinely leave out information which we assume the listeners will be able to fill in from their world knowledge. Any linguistic utterance thus only alludes to one of many possible subsets of a concept's features, one of many possible gestalt constructions. A theory of how speakers go from conceptualisation to articulation was fully worked out by Levelt (1989). The examples of systematic patterns in metaphor we saw earlier are examples of patterns in gestalt construction. Conventionalisation interacts with this underspecification in an interesting way: frequent use of one gestalt construction may lead speakers to assume that the conceptual features that the construction makes salient are more characteristic and defining. As we saw in the gun cleaning example, the English and Shawnee gestalt construction are equally apt to specify the place and manner of cleaning, so it would be questionable for,

say, English speakers to intuit that the Shawnee gestalt construction was less plausible than the English construction, and vice versa for Shawnee speakers. But conventionalisation may make speakers believe that the conventional gestalt construction was more plausible, more apt, to allude to the fully fleshed out conceptual knowledge than unconventional constructions. In the experiments of this thesis, participants will see unfamiliar (L2) metaphoric gestalt constructions but, like the English and Shawnee gunmen, they should have very similar conceptual knowledge pertaining to these constructions. We are thus curious how much speakers are influenced by conventions of metaphoric gestalt construction when they actually have very similar conceptual knowledge. If CMT's position is correct, then metaphoric linguistic expression would be a reflection of cognition and then metaphor-like similarity matching would also be fundamental to analogical reasoning.

Reddy (1979) constructs a thought experiment to fathom the scope of the Sapir-Whorf hypothesis: suppose there are people living in two worlds, which are, in a lot of respects very similar. However there is a barrier between the two worlds such that the people living in either one cannot talk to each other directly face to face. They can, however, communicate by sending written messages back and forth. Essentially, the two worlds correspond to the minds of two speakers. Clearly, we cannot read each other's minds, but we can communicate through language by implying and inferring meaning. The thought experiment poses the question: How is it, or is it even possible, that we can come to have similar concepts in our minds if we can only exchange these concepts by imperfect means of implying and inferring? And the answer that the thought experiments seems to suggest is that maybe it is the need to communicate effectively. Effective communication is required, for instance, to coordinate ourselves while cooperating in a task. If the task requires coordination in order to be successfully completed, maybe this pressure is enough to motivate us to find ways to align our conceptualisation. This line of thought was taken up by David Lewis (1969, 1975, 1979, 1983, 1986) and Paul Grice (1975, 1978, 1989) who, in their work, made it the central axiom which motivated the argument. They argued that, at least most of the time, speakers communicate in order to share information rather than to deceive, distract or confuse. Other interlocutors, so the argument goes, are therefore primed to assume this is speakers' communicative intent in most discourses. This alignment of speakers' communicative intentions and listeners' assumptions *about* speakers' intentions is what facilitates effective communication. On the Lewisian-Gricean view, it is this need to cooperate effectively that creates the necessary pressure and therefore communication, so they assume, also had to be effective. On this view, speakers adhere to rhetoric and other linguistic conventions because listeners' knowledge of the very same conventions increases the chance of inferential alignment between interlocutors.

Lewis (1969), for instance, thought about the emergence of linguistic conventions in terms of a problem for contractualism. He insightfully pointed out that we cannot think of linguistic conventions, such as those relating to meaning and concepts, as contractual agreements, not even in a hypothetical sense, where somehow our ancestors agreed on a convention, which is then simply handed down from one generation to the next. Lewis saw that this did not explain how people would have ever converged on a common convention in the first place and he argued that what was needed to cause this convergence was the need to cooperate and coordinate ourselves, which, he supposed, was only possible if the same need shaped language. Grice (1975, 1978) codified this need in the Cooperative Principle, from which he then derived the principles he thought shaped effective communication: the maxim of Quality (truthfulness), Quantity

(limiting the amount of information communicated to the amount necessary, but not more), Relation (the requirement that the information provided be relevant to the discourse), and Manner (to communicate information in a way appropriate to the discourse). Since then, many have criticised the shortcomings of Grice's definition of these principles, but I shall omit them here since it is not my intention to provide a repair to the Gricean maxims in this thesis. I will instead jump to the most recent approach, Relevance Theory (Sperber & Wilson, 1986; Wilson & Sperber, 1993; Sperber & Wilson, 1995; Wilson & Sperber, 2002, 2004), which replaces the central principles with a *cognitive* notion of relevance of the communicated information to the discourse. Relevance Theory is arguably one of the most successful approaches to the problem of how we align our mental representations through implying and inferring (van der Henst, Carles & Sperber, 2002; van der Henst & Sperber, 2004).

I am not convinced that we can take cross-linguistic differences in ways of conceptualising, i.e. the linguistic ways of *expressing* this way of conceptualising, as *direct* evidence of conceptual differences. Therefore, a central point of this thesis is to determine *how*, *when*, and *why* this conclusion is justified. In other words, I understand our task here to be to distinguish cases where a cross-linguistic difference in the linguistic form *is* indicative of a difference between conceptual systems (true positives) and cases where a cross-linguistic difference in the form is *not* indicative of conceptual differences (false positives). Note, the reverse is also true: cases where we find the same linguistic form in two languages although they relate to different concepts (false negatives) and cases where the same linguistic expression in two languages truthfully indicates that speakers also conceptualise in the same way (true negatives). At the end of this chapter (Section 2.7) I provide cross-linguistic examples of non-correlation between linguistic expression and conception and at the end of the next chapter (Chapter 3, Section 3.4) I show similar cross-linguistic non-correlations for metaphoric proverbs with idiomatic meanings, proverbs such as those in the experiments of this thesis.

2.4 Concepts as abilities

The concepts as abilities view (Dummett, 1993; Bennett & Hacker, 2008; Kenny, 2010) offers an interesting position in debating the question: How is it that conceptual systems vary interpersonally and cross-culturally so much when the neural basis, at its core, in its fundamental mechanisms and biological foundation, is the same for everyone of us? The abilities view investigates the nature of concepts from a behavioural perspective, which makes it an excellent candidate for empirical validation. Abilities that are indicative of concepts include, for instance, the ability to discriminate correct and wrong referents of a concept, in the real world, or the ability to successfully complete a task that requires knowledge and understanding of a concept. However, the abilities view tells us little about the structure of the conceptual system which is why I do not consider it a full alternative to the mental representations and Fregean accounts.

The seeds of the abilities view are already in the representational terminology of perception and apperception because in order to distinguish apperception from sensory perception and imagination we need to be able to tell internal from external concepts, which in turn requires analytic proof combined with experimentation, as I describe it in the next paragraph. Of course, in these early representational definitions the respective authors did not have the same physiological evidence we have today to ground their theories of perception, but it is important to note that the developments in theories of concepts are closely correlated with developments

in our understanding of the sensory and perceptual system and our understanding of where the divide between conscious and unconscious perception lies. And the abilities view makes an important contribution by incorporating behavioural evidence into the study of concepts. Fundamental to the abilities view is the rationale that if one group of people can only successfully solve a particular problem (a task) if they have knowledge and an understanding of a particular concept and demonstrate they can solve the problem, then another group, who can also solve the problem, must also have that concept.

The abilities view tries to circumvent the controversial question what the mental content of concepts is by shifting the object of study and theorising from the mind to behaviour. The first issue that concerns me is that because of this shift in the object of study, the abilities view gets caught up in secondary questions concerning the nature of abilities and when and how they relate to concepts rather than the nature of concepts themselves. The abilities view amounts to saying that all concepts are reducible to abilities. Can that really be true or are there concepts which are not reducible to abilities? The answer to this question hinges on the scope of the definition of ability. If abilities are claimed to be indicative of concepts, then in order to be able to capture *all* concepts through behaviour, the definition needs to be wide enough to apply the label ability to all corresponding behaviours. However, we must ask whether such a widening of the notion of ability is adequate and I am concerned that it might not be.

Since the abilities view measures ability as behavioural performance, the most basic and natural notion of ‘ability’ would have to use as a starting point for a definition would have to be performance in problem solving. A narrow definition seems suitable where having a concept is necessary in order to complete a task. For instance, in order to solve for x in the mathematical problem $1 + x = 3$, the individual would need to have the concept of NATURAL NUMBER, or at least NUMBER, and a concept of COUNTING. A narrow definition of ability might also be the ability to discriminate correct and wrong referents. For example, demonstrating the possession of the concept DRAGON might be taken to be the ability to pick the corresponding picture from a series of pictures. Of course, we would be inclined to say that a participant who picks the picture of the dragon does so because he or she knows what a dragon is, but this task could also be solved in other ways. For instance, as long as one can identify the creatures in the other pictures, one could solve the task by eliminating the only picture for which one does not have a concept. This demonstrates that behavioural observation is ambiguous and subjective. Consider also that another way to solve the dragon task: chance capitalisation. Participants who do not have the concept of DRAGON and therefore would not be able to identify the corresponding picture, might also not know what the other pictures show. They may nevertheless successfully pick the picture of the dragon simply by guessing. In order to ensure the reliability of such performance tests, participants’ motives for a particular choice are just as important as the correctness of that choice.

We do not have to go so far as to accuse participants of guessing and cheating. Consider a case where a single ability requires multiple concepts. Take the mathematical problem $1 + x = 3$, given earlier, again. I said that in order to answer correctly that $x = 2$, we need both a concept of NUMBER (maybe more specifically NATURAL NUMBER) and of COUNTING (with ADDITION, SUBTRACTION, and the other notions of arithmetic as sub-concepts). We might imagine a task that tests possession of the concept NUMBER. For instance, being presented with a series of symbols, only one of which is a number, while the others could be orthographic letters or pictographic symbols, perhaps. But how would we design a task

that tests for the concept of COUNTING *without* requiring the concept of NUMBER to be employed by participants in solving that task? I do not think such a task can be designed. If this is so, then there are concepts that cannot be tested because there is no single ability which correlates unambiguously with them. In the case of the concept of COUNTING we see that because it makes reference to the concept of NUMBER, we could only make inferences about people's ability to count (and them having the concept COUNTING) *indirectly*. So while non-correlations caused by guessing and cheating might be eliminated by improving test formats, non-correlations such as in the case of COUNTING cannot be eliminated. The question here is whether we can test for abilities that are in some sense parasitic on other abilities. If we cannot test for them separately, we can at least distinguish in principle between performance on tasks that require both COUNTING and NUMBER and performance on tasks that require only NUMBER.

What concerns me in all of these examples is that by shifting the object of study and theorising from concepts to abilities, the abilities view has become a theory that seeks out systematic patterns in the correlation and non-correlation between concepts and abilities rather than the nature of concepts themselves. By diverting our attention away from the nature of concepts themselves the abilities view undermines its own explanatory power. Performance and the mental content of abilities is already one of the objects of study in Embodied Cognition. As I said in the beginning of this section, the original motivation for the abilities view is to circumvent the controversial issue of the mental content of concepts by shifting the object of study from the mind to observable behaviour, abilities. However, Embodied Cognition studies the mental representations of sensorimotor procedural knowledge required for specific abilities and, as I said in Section (2.2), in the framework of Embodied Cognition, the position is that the content of at least some concepts is a function of its compositional structure and the contents of its constituents, and as the experiments by Eleanor Rosch suggest, means that for some concepts this already includes procedural knowledge as constituent content. Embodied Cognition therefore already studies the mental content and representation of abilities. The abilities view, on the other hand, makes no claims about the mental content of abilities, but instead focuses on performance. It thus would seem to me that the framework of Embodied Cognition already subsumes the abilities view.

In the experiments of this thesis I employ the abilities view to ground the way I operationalise the notion of metaphor plausibility. Plausibility judgements can be understood as the intuitions that are the output of the ability to evaluate the plausibility of the analogy implied between a metaphor's source and target concepts. It is the ability to gauge the amount of justification there is for drawing inferences necessary to meet the conditions imposed by the implied analogy. The more justification there is, the more plausible the metaphoric analogy should be. The fewer conditions have to be satisfied in order to justify the implied analogy, the easier it should be for speakers to judge metaphors' plausibility and the more plausible these metaphors should be.

2.5 Concepts as Fregean senses

Gottlob Frege (1892a,b,c) noticed that expressions in natural language behave like functions. Functions return output values based on specific input values. For instance, we can think of the word *father* in expressions such as *my father* or *the father of x* as a function that returns a particular person as output for the person that *my* and *of* point to as input. The function,

the natural language expression, imposes restrictions on the input values. For instance, only persons but not inanimate objects can serve as the input of the function *father of*. And the function imposes restrictions on potential outputs. The output of the function *father of*, for instance, can only be a man, not a woman. The function *father of* returns a different output for a different input: for instance, my father is a different person from your father. The corresponding concept FATHER, therefore, has many potential external referents, many fathers. Similarly, there are many external referents for the concept BOOK. We could therefore say that the concept BOOK is the mental representation of the class of real world entities (objects) that it refers to. In Section (2.2), I mentioned prototype theory, so we might assume that the mental representation involves some form of abstraction where the mental entities in this class are not representations of every and all real world entities that a speaker has perceptually experienced, but a prototypical idea of them that contains only their essential, distilled characteristics and defining properties. On Frege's view, natural language is a system of sense relations that allows us to express particular ways of referring. Referring to a particular object in the real world as the reference of a sense relation is only adequate, i.e. 'true', *with* the corresponding concept in mind. We can therefore think of a concept as a truth-value function which only yields 'true' if the correct sense is used to refer to its object. In other words, the truth-value function only yields 'true' if a natural language expression is used in the *right* sense to refer to the object (or set of objects) of that concept. On this view, the meaning of an expression in natural language is the particular relationship in which it stands to that which it refers to. Sense relations are ways of referring. Natural language, then, is a propositional system of labels for sense relation functions. Therefore, on Frege's view natural language expressions only have meaning within the compositional system of sense relations (Frege, 1892c, 1879). Functions that always yield truth values instead of abstract mental entities are concepts. In order to distinguish Frege's notion of concepts as truth-value functions (as opposed to non-truth-value functions) from the notions of concept in the mental representations and abilities view, I will refer to Frege's notion as Fregean concepts.

For *father* it seems there is only one sense and it is easy to take this single sense to be the concept itself, but many words have multiple senses and the different senses can be thought of as different aspects of a concept. In cases where natural language expressions have multiple senses, we thus see that senses are *distinct* from concepts. If a natural language expression has multiple senses, the specific sense in which it is used in a particular utterance can only be determined within the propositional system of a language, the discourse context, and in relation to the other senses of its own and other concepts. For instance, the word *time* can be used in the sense of 'a point in time', 'a period of time', and so on, all of which we could say are part of the concept of TIME, but notice that *time* can also be used in the sense of 'musical metre', which is a more specific notion of a periodic interval of time because it has a cyclicity to it that cannot necessarily be said to be part of the meaning of *time* in the sense of a 'period of time'. Hence, *time* used in the sense of 'musical metre' conveys a linguistic meaning that is distinct from its other senses. Whether or not *time* means 'musical metre' depends on the context of the utterance and discourse in which it is used: in the utterance *the piece is set in 3/4 time*, *time* can hardly mean anything other than 'musical metre', whereas in *he is doing time in the county jail*, it is highly unlikely that *time* means 'musical metre'.

Names are prototypical Fregean concepts: I can only refer to a person using a particular name, if it is indeed that person's name. At the same time, if you and I were in a conversation

and we referred to someone who is not present by their name, say *Jill*, we need to know who it is we are talking about in order to evaluate the meaning of statements such as *Jill couldn't make it to the meeting yesterday*, for instance. I can only determine whether the statement is true if I know who we are referring to, who Jill is, and whether or not she actually was at the meeting. I can have a concept of a person, an abstract idea of who *Jill* is, memories of her, a set of defining characteristics, character traits, an idea of what she looks like, and so on, but notice that the concept of JILL is different from concepts such as BOOK or FATHER. Suppose I knew several women named *Jill*. I would have a unique concept JILL for each of them ($Jill_1, Jill_2, \dots, Jill_n$) because they are all different people, uniquely defined by their characteristics, whereas I can say that there are characteristics that all books and fathers share that constitute the core of the concepts BOOK and FATHER. In this sense, names are arbitrary; hence their sense relations are arbitrary: I may know several people named *Jill*, but there is nothing, no JILLNESS, which requires that, by virtue of having the name *Jill*, implies that they share certain characteristics. However, it would be wrong to conclude from this that the sense of a name would be semantically vacuous. Even though there is no JILLNESS that could be said to be the semantic content of the sense of the name *Jill*, names are not merely indexical but have non-vacuous semantic content whenever one person has multiple names: for instance, the person named *Eric Blair* is also known under the name *George Orwell*. Different names for the same person constitute different Fregean sense relations, different perspectives on that person, and each unique perspective has its own distinct meaning. At the same time, all of these unique perspectives could be said to be part of the concept we have of that person. If I know that a person has two names, *Eric Blair* and *George Orwell*, or a nickname, for instance, I would find it equally adequate to refer to that person by either name; the two senses are two ways of connecting our concept of that person to its referent in the external world, but at the same time each sense focuses on particular aspects of the concept. Part of the sense of *Eric Blair* is that it is that person's given name and part the sense of *George Orwell* is that it is a pseudonym. That is why different names for the same person lack the tautological triviality of repeating the same name. We can say *George Orwell is Eric Blair* and it does not sound redundant, but conveys the intent to communicate that which differentiates the perspectives of the two names, that the name *George Orwell* is the pseudonym which the person whose given name is *Eric Blair* uses. Saying *Eric Blair is Eric Blair*, on the other hand, does sound redundant and trivial. Similarly, *Kal-El*, *Clark Kent*, and *Superman* are three different names, or Fregean sense relations, for the same person. They are distinct meanings because each expresses a different aspect of or perspective on the person. *Kal-El* expresses that this is the person's birth name, *Clark Kent* expresses that this is the person's human name given to him by his human foster parents, and *Superman* is the superhero alias chosen by that person to hide his other two identities. Our concept of the person might include the knowledge that all three names refer to the same person. Some other person, for instance Lois Lane, might not be aware of this. She would, nevertheless, treat all three names as distinct linguistic sense relations, just like we would. We can say *Clark Kent is Superman* to reveal Superman's true but secret identity, while saying *Superman is Superman* or *Clark Kent is Clark Kent* may again be trivial, or only pragmatically enriched by context. For instance, suppose Lois Lane has been speculating about Superman's true identity and Clark Kent tries to dissuade her by saying *Superman is Superman* to mean something along the lines of *just let it go, we'll never know who he really is*. This inference is the result of Clark Kent implying that *Superman is Superman* is semantically vacuous. Also note

this applies to metonymy, too: I can say *Stephen Hawking is the new Albert Einstein* to convey something like *he is a similar genius of physics*. Hence, reusing the same Fregean concept for the same external referent conveys no additional meaning (beyond pragmatic enrichment from context), while different Fregean concepts applied to the same external referent or the same Fregean concept applied to different referents do convey additional meaning. *George Orwell* and *Eric Blair* must be distinct aspects of meaning and *Clark Kent* must have a different meaning than *Superman* and *Kal-El*.

The extension of a Fregean concept is a *set* of external referents. This includes the possibility that the set may be empty, { }, which naturally lends itself to explaining concepts of mythical creatures or fictional entities such as *dragons*, *hobbits*, or *unicorns*. In Carnap's terminology (Carnap, 1937, 1942), intensions determine extensions. Extensions are constituents of the real world. Intensions are the cognitive correlate of extensions. We know *dragons*, *hobbits*, *unicorns* and the like do not exist in the real world. Hence, there are no creatures that could be identified as their external referents, their extensions. The extension of the concept FATHER is the set of all fathers; for the concept TREE it is the set of all instances of trees, and for names there is always and only one referent for every unique concept of a person. We might say that *dragons*, *hobbits*, and *unicorns* exist in works of fiction and that the mental representations of the corresponding concepts of DRAGON, HOBBIT, and UNICORN are rooted in the cultural descriptions and depictions of them. So although they are fictional, we could argue that their characterisation in these artworks is what serves as the source from which the mental content and representation of their concepts derives. In that case, the set of referents that is their extension would not be empty but be these descriptions and depictions. At the end of Section (2.2), we already encountered the challenges that fictionality poses to accounts of concepts. I shall not try to resolve these challenges because in the metaphor material of the experiments in this thesis, there are no fictional concepts. Rather, the important observation at this point of our discussion is that Frege supplies us with the terminology to characterise these challenges. A much more pressing issue for this thesis is that in preparing the experimental material, language-specific metaphors will have to be translated. In their translation, we inevitably have to decide on a choice of words, where each word comes with language-specific Fregean senses. We therefore need to find a way to ensure that metaphors' source and target concepts are preserved in translation.

In this thesis (see Chapter 6, Section 6.2), I use the notion of Fregean senses to establish the degree to which two concepts are similar: the degree to which two concepts are similar is taken to be proportional to the ratio between those Fregean senses that are common to both concepts and those that are not. We can use this notion of conceptual similarity to compare concepts across languages. Take the concept BEAUTY, for instance. According to WordNet (2010), the word *beauty* has three senses in English: (1) the qualities that give pleasure to the senses, (2) a very attractive or seductive looking person (*smasher*, *stunner*, *knockout*, *beauty*, *ravisher*, *sweetheart*, *peach*, *lulu*, *looker*), and (3) an outstanding example of its kind (*his roses were beauties*, *when I make a mistake it's a beaut*). According to the Duden (2013), the official German dictionary, *Schönheit* has these three senses in German: (1) the state of being beautiful (*attractiveness*, *vibrance*, *charisma*) (2) the qualities of a thing that make it beautiful (*elegance*, *chic*, *style*), (3) an attractive or beautiful person. English and German speakers share at least two of these three senses. We could therefore conclude that when the words *beauty* and *Schönheit* appear in the linguistic form of metaphoric proverbs, English and German speakers

are likely to have similar concepts of BEAUTY.

According to WordNet, *time* has the following ten senses in English: (1) an instance or single occasion for some event (*this time he succeeded, he called four times, he could do ten at a clip*), (2) an indefinite period (usually marked by specific attributes or activities, *he waited a long time, the time of year for planting, he was a great actor in his time*), (3) a period of time considered as a resource under your control and sufficient to accomplish something (*take time to smell the roses, I didn't have time to finish, it took more than half my time*), (4) a suitable moment (*it is time to go*), (5) the continuum of experience in which events pass from the future through the present to the past, (6) the time as given by a clock (*do you know what time it is?, the time is 10 o'clock*), (7) the fourth dimension/coordinate that is required (along with three spatial dimensions) to specify a physical event, (8) a person's experience on a particular occasion (*he had a time holding back the tears, they had a good time together*), (9) rhythm as given by division into parts of equal duration, and (10) the period of time a prisoner is imprisoned (*he served a prison term of 15 months, his sentence was 5 to 10 years, he is doing time in the county jail*).

According to the Duden, *Zeit* has the following ten senses in German: (1) a single occasion or limited period of time, (2) an indefinite period of time, (3) the passage of time as a series of moments, hours, days, weeks, and years, (4) a period of time under your control, (5) the time as given by a clock. The first five German senses are quite similar to the senses of *time* in English. The sense of *time* as the fourth dimension is usually not found in a German dictionary because it would be considered a scientific definition whereas dictionary is written with a non-scientific readership in mind. The Duden also lists this particular use of *Zeit*: (6) the official time within a particular timezone; surely English speakers would agree with this sense if *time* was used this way in English. However, the last four senses listed for *Zeit* in German are distinct: (7) period of time of historical or personal significance, (8) recorded duration of an athletic achievement (sports), (9) duration of a match (sports), (10) in linguistics, tense inflections. English speakers would probably find it more natural to refer to a period of time of historical significance using the word *age*, as in *the age of the dinosaurs*, but note that *time* is just as acceptable, e.g. *in ancient times*. Hence, English and German differ with respect to the sense relation, the word in natural language, which they use to express this aspect of the concept TIME. Similarly, English speaker would not use the word *time* in quite the same utterance context to refer to the duration of a match: whereas *Spielzeit*, literally *game time*, can refer both to the time the match starts and the duration of the match itself, English speakers might be more inclined to think it meant the time the match starts, whereas German speakers more commonly use it to refer to the duration of the game. Hence, although this is not really a conceptual difference, it certainly affects which inferences speakers are more likely to draw in comprehension. In English language education, the use of *time* for tense has fallen out of style, but it is still quite common in German education to refer to tense inflections as *Zeitformen* (literally *forms*, meaning inflectional forms, *of time*). The word *Zeit* is never used to mean musical metre in German. Rather, German uses the word *Takt*, which is akin to the use of *beat* for musical metre in English. German speakers use *Zeit* colloquially to refer to a prison sentence, but its use is much less common than in English.

Notice that although English and German speakers differ in their use of the words *time* and *Zeit*, it is not the case that English and German speakers had radically different notions of the concept TIME. However, because metaphors will have to be translated so that they can

be presented to English and German experimental participants in their native language, we inevitably have to decide which words, and thereby Fregean senses, we should choose in the translation of metaphors' source and target concepts. In cases where we have a choice between different words, such as *Zeit* versus *Takt* for the sense of 'musical metre', we should choose those words in translation that minimise cross-linguistic differences in Fregean senses. However, the takeaway message is that whenever we choose one natural language expression over another, it comes with all of its language-specific Fregean senses and not just those that correspond to the expression in the other language. Cross-linguistic mismatches in Fregean senses are mismatches in linguistic expression, vocabulary, and word choice, and we need to carefully decide when they correlate with conceptual differences. At the end of this chapter, in Section (2.7), I provide more examples of cross-linguistic mismatches between concepts and Fregean senses, particularly cases where cross-linguistic differences in senses might mislead us to think conceptual differences exist where there are none. Since I wish to gauge cross-linguistic metaphor communicability in the experiments of this thesis by maximising the cross-linguistic similarity of metaphors' source and target concepts, such *extreme* cases of concept-sense non-correlation are of particular interest in order to identify to what extent cross-linguistic conceptual similarity is attainable. In the next section, we look at linguistic meaning more closely. Frege's contribution to this discussion is that he showed that concepts and senses are distinct aspects of linguistic meaning.

2.6 Linguistic meaning

The nature of linguistic meaning is controversial in linguistic theorising and follows either (1) the formal-semantic tradition that sees semantics as the part of the grammar that takes syntactic units as input and pairs them up with meanings constrained by the truth conditions of linguistic utterances or (2) the non-formal pragmatic tradition that argues that semantics, in the way that the formal-semantic tradition envisions it, cannot produce the full meaning of utterances that speakers intend to imply and that listeners infer. Advocates of the non-formal pragmatic tradition usually propose that the gap between sentence/utterance meaning and speaker meaning, the full meaning implied and inferred, is to be bridged by employing pragmatics in order to draw the necessary inferences. The position I take in this thesis is that linguistic meaning is speaker meaning, not sentence/utterance meaning. I thus take linguistic meaning to be meaning in context, including pragmatics, speaker intent and what she meant to imply, and what other interlocutors infer. The position I adopt here is thus in line with what CMT, RT and GS take meaning to be. I wish to investigate their predictions concerning the nature of metaphoric idiomatic meanings. It thus makes sense to adopt their notion of linguistic meanings.

Linguistic meaning consists of three components: (1) a concept, (2) Fregean sense relations, (3) reference/external referent/extension. These three components of meaning were recognised by de Saussure (1993, 1916), albeit he used the term *signifié* for concept, *signifié*r for sense relations, and *objet* for the external referent. This mirrors Frege's distinction between (1) *Konzept* (concept), (2) *Sinn* (sense or sense relation), (3a) *Bedeutung* (reference) and (3b) *Gegenstand* (object, the external referent, external from the mind, in the real world).

The notion of 'concept' in linguistic meaning is defined more narrowly than the one commonly held by the mental representations view, Embodied Cognition, CMT, and the abilities view. In linguistic meaning, a concept is thought of as distinct from Fregean sense relations. At

Table 2.1: Eight types of linguistic symbols based on the three binary features *concept*, *sense relations*, and *external referent*. Notation: A linguistic symbol has [+] or does not have [-] a concept (C), sense relations (S), or an external referent (R).

	C	S	R
Unnamed, abstract concept	+	-	-
Fictional (e.g. <i>dragon</i> , <i>unicorn</i> , <i>hobbit</i>)	+	+	-
Typical linguistic symbol	+	+	+
Inconceivable but named thing. Possible?	-	+	+
Real but undiscovered thing	-	-	+
Real and known thing with no word for it	+	-	+
Nonsense symbol (e.g. <i>wug</i>)	-	+	-
Nothing	-	-	-

the same time, the notion of ‘concept’ in linguistic meaning is akin to mental representations. In this thesis, I make no claims about concepts’ mental representation. All that is necessary for the arguments in this thesis is that we agree that we can assume that two speakers can gauge and assess whether they have similar concepts in mind for a particular word in natural language (*comparability of conceptual similarity*). The notion of similarity employed here will be specified in Section 2.7. A linguistic symbol consists of (1) a form (e.g. a word, phrase, sentence, etc.), (2) a meaning that relates the linguistic form to a referent in the external world, and (3) an association between the form and meaning, which may be more or less arbitrary. In linguistic meaning, Fregean sense relations are the specific linguistic forms and sense relations are thus used in distinction to concepts proper. The performance perspective of the abilities view is inherent in the object of linguistic study: language as a behaviour.

Each of the three components of linguistic meaning (concepts, sense relations, and referents) might be given independent of the other two for a particular linguistic symbol. For instance, we might have a concept with no word for it in natural language and this concept might be abstract because it does not refer to anything in the external world. In the case of fictional entities such as *dragons*, *unicorns*, and *hobbits*, we can have a concept of what a dragon, a unicorn, or a hobbit is and there are words in natural language to express these concepts, i.e. there are Fregean sense relations for them, yet there are no creatures in the external world that could be said to be the referents of these linguistic symbols. Many typical linguistic symbols, such as nouns, adjectives, and lexical verbs, have all three components of meaning: a concept, Fregean sense relations, and an external referent. The noun *book* refers to one instance of the class of entities denoted by the concept of BOOK that is expressed through the sense relation *book* in English, *Buch* in German, *livre* in French, and so on. Proper nouns are special because sometimes there are multiple names for the same person. For instance the person whose birth name is Eric Blair uses the name George Orwell as a writer. Hence there are two Fregean sense relations, *Eric Blair* and *George Orwell*, which mediate between the concept we might have of that person and the real person. Similarly, *Kal-El*, *Clark Kent*, and *Superman* are three different names, or Fregean sense relations, for the same person. They are distinct meanings because each expresses a different aspect or perspective on the person. The adjective *green* is the sense relation used in English to convey the concept GREEN (or GREENNESS) which

takes all entities in the external world that it applies to (i.e. all green things) as its potential referents. The lexical verb *to run* is the sense relation used in English to convey the concept RUN and it applies to all acts of running in the external world as potential referents. While nouns, adjective, and lexical verbs are cases where it is easy to characterise reference, this might be harder for other part-of-speech classes. For instance, what is the external referent of the preposition *up*? On the conceptual side we need to be aware that the sense relation *up* can refer to different concepts: a directional concept UP as in *walking up the stairs*, a locational UP as in *I live up the road*, an emotional UP as in *I'm feeling pretty up at the moment*, and a concept corresponding to idiomatic uses of *up* as in *to look up an article*. Due to the fact that all of the source and target concepts in the metaphors in the experiments of this thesis are typical linguistic symbols with a concept, sense relations, and referents, we shall not concern ourselves with part-of-speech classes where the characteristics of reference are problematic and controversial. An entity in the external world that has not yet been discovered can as a consequence not yet have an established concept and sense relation. There might also be cases where we might have a concept that relates to something that exists in the external world, yet where no conventionally accepted sense relations have been established. For instance the feeling of shopping bags cutting into one's hands. The concept exists because it can be explained, for instance, through the short description I just gave. A sense relation need not be a single word in natural language, it could also be a phrase or sentence, but I would argue that for this example concept no sense relation has been *conventionally* established. The linguistic symbol *wug* has no meaning in English; it is a nonsense symbol. It can be taken to apply to anything or nothing. Hence, *wug* is a sense relation without a concept to refer to and without an external referent; or conversely it could apply to any concept and take any entity, property, or action in the external world as referent. The potential configurations of linguistic meanings are summarised in Table (2.1). All metaphor source and target concepts in this thesis are typical linguistic symbols in that they have concepts, sense relations, and referents, i.e. they are of type [C+,S+,R+]. None are fictional, none are unnamed, none are undiscovered or inconceivable; they are of a very general sort so that it should be safe to assume that all humans would be concerned with them. They speak to our human nature in general: e.g. love, friendship, morality, a person's character and other qualities, and one's outlook on life. Speakers of languages with a shared cultural background and historic intellectual interaction have a higher chance of relating to these concepts in similar ways.

The metaphoric proverbs in the experiments of this thesis have conventional idiomatic meanings. An idiomatic meaning is a linguistic meaning that cannot be inferred compositionally from the linguistic form of the utterance but is given by linguistic convention within a particular speech community. Idiomatic meanings are thus conventional (salient) linguistic meanings. An idiom is a linguistic utterance where the association between its form and meaning is arbitrary because the meaning cannot be inferred compositionally from its constituents. Proverbs are idioms because the association between linguistic form and meaning is arbitrary and cannot be inferred compositionally on the basis of the conceptual and pragmatic knowledge associated with its formal constituents. A proverb is a phrasal or sentential idiom. Its linguistic form is either that of a phrase or a sentence. I understand 'linguistic convention' here in the sense of Lewis (1969) and Grice (1975, 1978): a linguistic convention is an emergent regularity in linguistic expression that emerges out of the need for effective communication. Linguistic conventions might reflect conception, but they need not (see the Sapir/Whorf Hypothesis in

Section 2.3). By assuming that systematic patterns in linguistic metaphor reflect salient inferences (conceptual metaphors, in CMT terms), CMT takes them to reflect conception rather *directly*. RT and GS, on the other hand, take patterns in linguistic metaphor to be rather *indirect* evidence of conception, *if at all*, but rather of linguistic convention. By comparing the predictions concerning metaphoric idiomatic meanings made by CMT, on the one hand, and RT and GS, on the other, this thesis investigates how much insight they give us into speakers' ways of conceptualising.

2.7 Concepts from a cross-linguistic perspective

In Section (2.5), we saw examples where there are multiple Fregean senses (linguistic forms) for the same concept. For instance, *Clark Kent*, *Kal-El*, and *Superman* are multiple names (Fregean senses) for the same concept of a person. Similarly, our concept of GEORGE ORWELL may include the knowledge that *George Orwell* is the name that the person whose birth name is *Eric Blair* goes by as a professional writer. Our concept of the person thus has two Fregean senses, two meaning perspectives, in the same way that *Clark Kent*, *Kal-El*, and *Superman* convey different meaning perspectives on our concept of the person. These examples show that the correlation between concepts and their linguistic form (Fregean senses) is not always one-to-one. We therefore need to be wary that non-correlations can and do occur. This issue becomes even more pressing when we attempt to gauge the similarity of concepts cross-linguistically. In this thesis, we are interested in the nature of inferential comprehension of idiomatic meanings of metaphoric proverbs by native and non-native speakers. Idiomatic meanings are conventionalised default interpretations that act as inferential shortcuts. CMT proposes that what is conventionalised in metaphoric idiomatic meanings are *conceptual* associations called conceptual metaphors; RT and GS propose that what is conventionalised are *non-conceptual* linguistic conventions of how complex metaphoric meanings and their imagery are expressed (in addition to conventionalisation of concepts themselves). A metaphor consists of two concepts, a source and a target concept. If, as CMT claims, speakers rely on their knowledge of concepts and conceptual metaphors to infer the metaphoric analogy implied between the source and target concept, then for two languages where, for the metaphor in question, the source and target concept and the repertoire of potentially inferentially relevant conceptual metaphors are cross-linguistically similar, then native and non-native speakers should infer similar linguistic meanings for this metaphor. If, as RT and GS argue, cross-linguistically similar conceptual knowledge and similar mental imagery is not enough to maximise cross-linguistic metaphor communicability, but similar non-conceptual linguistic conventions are required *in addition* to conceptual similarity, then for two languages where a metaphor's source and target concept and relevant mental imagery and conceptual associations are cross-linguistically similar, we should find lowered cross-linguistic metaphor communicability and the magnitude of lowering should correspond to the size of the effect that non-conceptual linguistic conventions have on metaphor comprehension. In order to experimentally test the predictions made by CMT, on the one hand, and RT and GS, on the other, we therefore need to pick linguistic metaphors with language-specific idiomatic meanings and where the source and target concepts of these metaphors, as well as relevant mental imagery and conceptual associations, are maximally similar for two languages.

In order to maximise conceptual similarity we must first be able to gauge it. In the way

of that stands the following unresolved and controversial issue of linguistic theorising: No two interlocutors may have exactly the same concept in mind for the same expression in natural language yet inferential communication is possible and successful when using that expression. Colours are an excellent example to illustrate this issue. Light that is visible to the human eye makes up only a very small part of the electromagnetic spectrum. Light consists of particles called photons. Photons have different energy states which are commonly given as their wavelength. We commonly speak of the wavelength of light, but in fact, the light of the sun, for instance, consists of photons of many different wavelengths that span the wavelengths that are visible to the human eye but also contain ultraviolet and infrared wavelengths that are invisible to us. The retina in the eye contains light-sensitive cone-shaped cells which are specialised to react to particular wavelengths of visible light. These cone cells are the first source of perceptual variation: the average human has three to four types of cone cells. However, some individuals only have two types and others can have up to sixteen types. As a result, individuals with different numbers and types of cone cells also differ in how many hues of colour they can perceive and distinguish. Hence, there is a great deal of perceptual variation, even among those who, medically speaking, have ‘normal’ vision. The cone cells send electric signals to the brain. The signals are stronger for a particular cone type, the more photons of that wavelength the light contains. What the visual nerve thus communicates to the brain are wavelength-dependent light intensities. The impression of colour itself is created in the mind. So not only is there biological variation in the number and type of cone cells, but there is also interpersonal variation in how the individual brain has learned to interpret the visual signal. Hence, even two individuals with the same number and type of cone cells may have brains that interpret the visual signal differently, thus leading to different impressions of colour for the same physical wavelengths of light. In colour blindness these interpersonal differences become apparent because people with the condition fail to see colour differences that people with normal vision can see. Colour blindness is diagnosed using the Ishihara color test: people are shown patterns of coloured spots. Hidden in the spot patterns are numerals or letters, for instance, a green number hidden among red spots. This is a test of *communicability*: those with normal vision will be able to make out the green spots among the red ones and will be able to recognise the number character that the green spots form. Those with red/green blindness, however, will fail to see the number. The spot pattern will thus communicate something to people with normal vision that people with red/green blindness will be unable to infer. But it also means that people with colour blindness do not have the same colour concepts as those with normal vision: in the same way that one cannot communicate to a blind person what the sensations of colours are like or why RED is warm and BLUE is cold, one cannot communicate to a person who is red/green blind what the difference between RED and GREEN is. Individuals with normal vision might agree that when one speaker uses the word *red* to refer to light of a particular wavelength that another speaker also calls it *red*. However, we cannot say if they also share the same mental sensation of colour which they have conventionally come to associate with their own concept of RED.

There is neurophysiological evidence for interpersonal variation in the mental representation of episodic memory which is rich in individual concepts, their functional distribution and topological brain regions (Thompson-Schill, Braver & Jonides, 2005; Miller & Van Horn, 2007; Van Horn, Grafton & Miller, 2008). These interpersonal differences are stable over time (Miller, Van Horn, Wolford, Handy, Valsangkar-Smyth, Inati, Grafton & Gazzaniga, 2002). Within the same person, however, mental representations are stable for the same task and over long pe-

riods of time (McGonigle, Howseman, Athwal, Friston, Frackowiak & Holmes, 2000). Despite interpersonal variation in the location of individual concepts, the mental content of these concepts might nevertheless be quite similar. I therefore take conceptual similarity to be *gradient*. If no two speakers can have exactly the same concept in mind for an expression in natural language, the question then becomes how dissimilar can the concepts in speakers' minds be before communication is impeded? This question certainly has no general answer, but can only be answered for each concept individually. Nevertheless, we can be more specific about the restrictions that communicability imposes on the amount of inter-speaker conceptual variation.

I would argue that we can say that two concepts are (cross-speaker or cross-linguistically) similar enough if the between-speaker variation in the mental content and representation of the concepts does not interfere with their inter-speaker communicability. In other words, I argue that we can say that for the purpose of the experiments in this thesis, where we are interested in metaphor *communicability*, communicability should also be our gauge of conceptual similarity. So while I think that two speakers, be they speakers of the same or two different languages, can never have exactly the same concept in mind for a particular linguistic form in natural language, we may say that the concepts that they do have in mind can be said to be 'similar enough' if and only if the inter-speaker variation in mental content and representation is smaller for the two concepts in question than between either of them and all the other concepts that the two speakers have in their mental lexicons. In other words, the speakers' concepts are 'similar enough' (approximately conceptually similar) if and only if the cross-speaker conceptual similarity between them is greater than for any other concepts that the speakers have. If conception is a matter of similarity matching, let us think of 'conceptual similarity' as gradient, from 'not similar' to 'partial matches' to 'similar.' So if, for a particular expression in natural language, the first speaker has concept A in mind and the second speaker has concept B in mind, then among the first speaker's concepts that come close to A and among the second speaker's concepts that come close to B there will inevitably be two which are more similar than the alternative. This does, however, not mean that speakers may not make use of the alternatives, it only means that 'approximate conceptual similarity' is a likelihood which is maximized for A and B. If speakers' ability to use a natural language expression to communicate their mental states, then the communicability of it should be maximised when 'conceptual similarity' is maximized. So even though speakers cannot use human language to cause others to have *exactly* the same mental states and sensations as they themselves (e.g. to make them 'see' colours the way that they 'see' them with their minds' eye), we can say that the mental states and sensations that they *do* cause in others are 'similar enough' if communication is unimpeded and without interference.

We can then think of the public lexicon of a language (in RT terms, Sperber & Wilson, 1998) as an average of the mental lexicons of its speakers. If we nevertheless also want to account for the fact that individual speakers of this language will have mental lexicons that deviate from the average public lexicon, then we might say that the public lexicons of two languages are significantly different if their sets of averaged concepts, averaged across concepts' inter-speaker approximate similarity, are more different than the amount by which the mental lexicons of its speakers deviate from the public lexicons within the two languages. We might therefore think of the cross-linguistic similarity of the conceptual systems of two languages as a matter of degree that can be expressed as the likelihood of the mental lexicons of speakers of the two languages to overlap and, although they may never be identical and/or overlap perfectly, we

Table 2.2: Cross-linguistic differences in Fregean sense relations as evidence for conceptual differences.

	Similar concept	Different concepts
Similar sense relations	<i>True Positives</i> <i>beauty,</i> <i>blindness</i>	<i>False Positives</i> egocentric versus geocentric <i>left/right</i>
Different sense relations	<i>False Negatives</i> <i>pig/pork, porc;</i> <i>Samstag, Sonnabend</i>	<i>True Negatives</i> <i>Umami</i> as a distinct taste concept

may nevertheless be able to gauge when they are *more* similar for two given languages than they are for two other languages.

The following cross-linguistic examples demonstrate that cases of miscommunication are good indicators of lowered communicability. The examples are examples of cross-linguistic correlation and non-correlation between concepts (narrow linguistic definition) and their language-specific linguistic form (Fregean sense relations).

Since our empirical evidence is ultimately grounded in linguistic expression, we must ask ourselves *what is the relationship between linguistic variation and conceptual variation*. Let me give you a couple of examples that show that this relationship is not always one-to-one. For a systematic overview it helps to think about this in terms of hypothesis testing where we ask *how often does a difference in linguistic expression correlate with a conceptual difference*. When we think of this question as an empirical test, then the test result (the assumed correlation) might be mistaken. Hence, some of the time, the test might be truthful (true positive and true negative correlation), but it could also be wrong (false positive and false negative conclusion that a correlation exists where there really is none), and we might find it helpful to represent this as in Table (2.2). Let me give you a couple of examples for true and false correlations.

Japanese has a word for a taste concept called *umami*. Historically, it seems that *umami* flavours have long been available in Western cuisine (e.g. parmesan), but were not recognised as such until the Japanese identification of the category. There is physiological evidence that the human tongue actually has special taste buds dedicated to tasting *umami*. Hence, in Fregean terminology, we know that *umami* is a concept external to our minds and rooted in biology, and it just so happens that Japanese identified it as a category. We are certain that Western cuisine had neither a word nor concept for this specific taste until it adopted the Japanese term *umami*. Another example is *bokeh* blur, also from Japanese, which found its way into the terminology of Western cinematography to describe the specific type of aesthetically pleasing focal blur associated with cameras typically used in film. The term was already used in the arts before the advent of modern cinema in Japan. Since cinematographers have a professional interest to improve the aesthetic quality of the film image itself, it makes sense to distinguish between focal blurs that are aesthetic, *bokeh*, and those that are not. So although we might say that the objective would be to have a sharp image, these cinematographers are very particular about what the out-of-focus portions of the background should look like and to describe them they use *bokeh* to mean those they find aesthetically pleasing, making all others *non-bokeh* (e.g.

they would say that a blur that is homogeneous, i.e. Gaussian, is not *bokeh*). Note that *bokeh* is not simply a synonym for *aesthetic*; cinematographers will use it to refer to a particular type of focal blur which has distinct visual characteristics. Notice that while *umami* turns out to be an external concept, *bokeh* is an aesthetic concept, hence internal.

Schadenfreude, in German, describes the concept of taking delight in someone else's misfortune. The compound noun literally translates to *misfortune-delight*. Nevertheless, just because German happens to have a word for this concept, the fact that English, for instance, does not have a word for it, does not mean that speakers of English would not be able to understand the concept and I would go so far as to argue that, in fact, most English speakers have this concept without having a conventional linguistic form of expressing it. Educated English speakers will use the German word for it. Hence, I would argue that English and German speakers share this concept and German speakers simply have agreed upon a conventional way of expressing it while (most) English speakers have not. Another similar example is English *pig* and *pork* versus French *porc*. Historically *pork* came as a French loan word into English, but while in English it refers only to the meat of the animal, in French *porc* is used to refer to both the animal and its meat. However, we know that French speakers conceptually make the same distinction between the animal and its meat, they distinguish the same two concepts PIG and PIG MEAT as English speakers. Similarly there are two distinct linguistic forms, *cow* and *beef* or *deer* and *venison* or *sheep* and *veal*, to refer to the concept of the animal and the concept of its meat in English, but only one linguistic form for both concepts in French. Hence, a conceptual distinction that we know speakers of both languages make, is only expressed linguistically by English speakers. Hence, this example shows that we cannot always assume that the existence of a conceptual distinction will necessarily correlate with that distinction being linguistically marked. Notice how this example is similar to the previous example of *Schadenfreude*, where both English and German speakers have the concept, but only German speakers have a distinctly, or marked, linguistic expression for it, in the same way that English speakers mark the distinction between *pig* and *pork* linguistically while French speakers do not.

Umami and *bokeh* are examples that are indicative of conceptual differences because before they became known outside of Japan, non-Japanese speakers had neither these concepts nor words for them. Hence, if as a non-Japanese speaker we came across these two words unfamiliar to us, we would be right to assume that they are names for unfamiliar concepts as well; hence, they are true positives. But if we were English speakers who came across the word *Schadenfreude*, it would be wrong of us to assume that it referred to an unfamiliar concept. And if we were speakers of French, when we see that English speakers use *pig* and *pork* where we use only one word, *porc*, it would be wrong of us to assume that English speakers must be making a conceptual distinction that we do not make. Hence, *Schadenfreude* and *pig/pork* are *false positives*.

In our hunt for conceptual differences across the world, we might also come across linguistic forms that seem familiar to the ones in our native language, but we realise that in this other language they are used very differently, so different that it leads us to conclude that there must be a conceptual difference between the two languages. For instance, some languages use an egocentric spatial reference system for the words *left* and *right* while others use a geocentric system. English, for instance, uses an egocentric reference system. Imagine you have just moved into a new apartment and you are in the process of arranging furniture. You and a friend have just manoeuvred a heavy sofa into position, but now another friend, who has been

standing back to make sure the sofa is in the right spot, says to you *No, more to the left*. And in most cases you would wonder *My left or your left* because *left* and *right* change from person to person and depending on where that person is facing, i.e. you are using an egocentric reference system. Speakers of Tzeltal (Li, Abarbanell, Gleitman & Papafragou, 2011), spoken by a group of Mayans living in the Tenejapa area of Chiapas (Mexico), in the same situation, however, would not be confused where to move the sofa because their language uses a geocentric reference system for spatial terms. Tzeltal uses *xin* (left) and *wa'el* (right) only for body parts. For directions speakers instead use geocentric directions such as north and south, i.e. *Move the sofa northward* would imply the same direction for everyone involved. But now imagine a scenario where both Tzeltal and English speakers were in the same situation together. How long would it take them to realise that they use different reference systems and come to a common understanding in which direction the sofa should be moved? Judging from how long it sometimes takes speakers of languages with egocentric reference systems to figure out which direction they are supposed to go even when they know they are using the *same* reference system, we can only assume that it would take them even longer if they did not know the miscommunication was not because of how they were facing but because they were using entirely different reference systems. Similarly, English speakers, not being accustomed to think in terms of geocentric directions when moving furniture, might be equally confused as Tzeltal speakers who are not accustomed to use *left* and *right* for body-external objects such as sofas. This is an example of *false negatives* because a conceptual difference that really exists between the two languages, is not marked by distinct linguistic forms. The point that interests me here is that it is the *miscommunication* that occurs between speakers of the two reference systems which tells us that they are using different reference systems *although* they use the same linguistic expressions in both systems. Hence, the experimental test, the methodology, that we would use to detect the conceptual difference is based on the concepts as abilities view.

We might assume that these vast conceptual differences as in the case of the egocentric versus geocentric spatial reference systems are more likely to occur between languages with little or no linguistic contact and exchange, but this conclusion can quickly be disproved. My native dialect of German, for instance, has another system of telling the time than standard High German. For instance, in my dialect I might say *viertel 3* (literally *quarter 3*) which would translate to *viertel nach 2* (*quarter past 2*) in High German, both of which, however, mean 2:15 am/pm. The time systems of dialect and standard both use *halb 3* (*half past 2* to mean 2:30 am/pm, but note that it literally says *half 3* in German). But then they disagree again in how to say 2:45 am/pm. My dialect uses *dreiviertel 3* (*three-quarters 3*) while in the standard it is expressed as *viertel vor 3* (*quarter to 3*). When we imagine a typical situation, setting a time for a meeting, we can see how the two systems can quickly lead to grave misunderstanding. When I say *wie wär's mit viertel 3* (*let's meet*, literally, *quarter 3*), the other person, who, let's assume, is not a speaker of my dialect, might think I simply omitted the *nach* (*past*, as in *quarter past*), so they think I meant 3:15 pm, or they might think I omitted a *vor* (*to*, as in *quarter to*), and think I meant 2:45 pm. The time I actually had in mind was 2:15 pm, so *regardless* of which assumption they make, they will show up between 30 minutes and an hour late. This example shows two things: (1) grave conceptual differences might be marked by small linguistic differences and (2) these small linguistic differences might thus lead to grave misunderstandings. So while, technically, this is an example of a *true positive* where a conceptual difference *is* correlated with distinct linguistic forms, the linguistic form is so small that even for non-dialect speakers of the same

language it might go unnoticed and they are likely to misunderstand which time I have in mind.

My native dialect also has another peculiar example: a different word for Saturday than in standard High German. This may lead to similar cross-dialect miscommunication as in the case of telling the time of day. The Christian week starts on Sunday, which makes Saturday the end of the week. Hence, if we think about the course of a week in conceptually similar terms to the way we think about the course of a day (i.e. via conceptual *analogy*), then the day starts at sunrise, hence *Sonntag* (*Sunday*, the first day of the Christian week, as in *at the break of daylight*) and ends in the evening. It therefore makes sense to refer to Saturday, the last day of the Christian week, as *Sonnabend* (*Sun-evening*). When I, as a dialect speaker, would use the dialect word for Saturday, *Sonnabend*, which literally translates to *Sun-evening*, non-dialect speakers might, in their logic, think *Sonnabend* (*Sun-evening*) sounds a lot more like *Sonntag* (*Sunday*) than *Samstag* (*Saturday*). So often they do not even consider *Saturday* as a viable option but dismiss it along with the other days of the week, which also do not sound anything like *Samstag* (*Saturday*) or *Sonntag* (*Sunday*), and think *Sonnabend* is simply a dialectal variation on *Sonntag* (*Sunday*). This interpretation makes sense conceptually, not just for non-dialect but dialect speakers as well, because in a non-religious context, the working week starts on Monday and ends on Sunday. So in their logic it might make sense to assume that *Sonnabend* (*Sun-evening*) refers to the end of the week, which is actually correct, but since they assume the reference system is the professional not the religious system, they assume since Sunday is the last day of the working week, Sun-evening must be Sunday. Note, however, that speakers of my dialect do not have religious connotations when they use *Sonnabend*, they simply use it in all contexts where non-dialect speakers would use *Samstag*. This is the reverse of the *pig/pork* example: While French speakers have a concept for the ANIMAL and its MEAT, they use the same word *porc* for both. Here we have an example where speakers of different dialects of German have different linguistic forms, *Sonnabend* (*Sun-evening*) and *Samstag* (*Saturday*), to refer to the same concept SATURDAY. It is therefore an example of *false positives*.

The vast number of instances, such as *beauty* or *blindness* that occur cross-linguistically, for instance in German as *Schönheit* and *Blindheit*, can be established on the basis of their Fregean senses in the respective languages. According to the Oxford English Dictionary, *beauty* has three senses: (1) the qualities that give pleasure to the senses, (2) a very attractive or seductive looking woman, and (3) an outstanding example of its kind. According to the Duden (2013), the official German dictionary, *Schönheit* has the same three senses as English *beauty*: (1) the quality of being beautiful, attractive, pleasurable, or charismatic, (2) a beautiful person, and (3) an outstanding example of its kind or an attractive feature of a thing. English and German speakers share at least two of these three senses. So while we would of course expect that individual speakers of English and German might have different ideas about what, say, makes a person attractive, we can at least say that *beauty* and *Schönheit* have the similar Fregean senses in the two languages and inter-speaker variation should operate within the bounds set by these senses. Following the rationale of the concepts as abilities view, we can say that if speakers of English and German seem to use the words *beauty* and *Schönheit* in similar ways (similar Fregean senses), then it is likely that they may also have a similar concept of BEAUTY (again, despite personal preferences of aesthetics, of course). Similarly, *blindness*, in English, has the senses: (1) an ability to see, (2) the ignorance or unwillingness to perceive or understand something, and (3) being wilfully unreasonable. In the Duden we find similar senses, albeit with different distinctions: (1) an inability to see, (2a) without reason, (2b) without critical thinking, (2c) not

prudent, being unreasonable according to common sense, (3) an unwillingness to be reasonable. We can thus say that English and German speakers should have similar mental representations for the concept of BLINDNESS. In Chapter 6, I apply this cross-linguistic analysis of the sense relations to all source and target concepts of the metaphors in the experiments of this thesis. In the experiments, we should try to maximise the cross-linguistic conceptual similarity of metaphors' source and target concepts. We should therefore choose metaphoric proverbs with idiomatic meanings and novel metaphors that resemble the form of the proverbs such that, cross-linguistically, there is a true correlation between the concepts and their Fregean sense relations in the two languages that the proverbs are taken from. In order to maximise the cross-linguistic similarity of the source and target concepts, we should choose cases of true positive concept-sense correlations. In Chapter 6, I propose that, in addition to picking concepts with true positive concept-sense correlations, we can quantify the degree of cross-linguistic conceptual similarity by taking the ratio of cross-linguistically similar and dissimilar Fregean senses as an index of conceptual similarity.

2.8 Summary

In this thesis, we will experimentally test CMT, RT and GS's views on metaphoric idiomatic meanings. Cognitively, a metaphor consists of (1) a source concept, (2) a target concept, and (3) an analogy that is implied to hold between the two concepts. At the beginning of this chapter I introduced the various views one might take as to what concepts are. These views fall roughly into three camps: (1) the view that concepts are mental representations, which are either rooted in perceptual experience, the experience of body-internal states, or which are abstractions of these perceptually embodied concepts or imaginary extrapolations of them (the view broadly referred to as Embodied Cognition). Of chief interest to this thesis is CMT, which takes the simplest concepts in the conceptual system to all be embodied concepts and all complex concepts to be rooted in primary embodied concepts. (2) The view that concepts are best understood through our ability to use them. If the completion of a task requires knowledge of a particular concept, then demonstrating the ability to complete the task should be indicative of having that concept. Metaphor comprehension is such a task: inferring the conditions under which the implied analogy between the source and target concept holds, requires knowledge of these concepts. If two speakers infer different meanings for a metaphor, then this should, at least partly, indicate that their conceptual knowledge and representation of the source and target concept differ. This is the view I employ to define in Chapter 4 how we should measure cross-linguistic metaphor communicability experimentally. (3) The Fregean view that concepts are abstract symbols whose meaning is only inferable within a propositional language. We will use Fregean senses to gauge conceptual cross-linguistic differences between the English and German translation of the experimental material.

CMT sees the meaning of a metaphor, the output of the inferential comprehension process, as a complex concept. On this view, the mental content and representation of a complex concept is a function of its compositional structure and the contents of its constituent concepts. The source and target concept are the constituents of the complex metaphor concept and the pragmatically enriched implied analogy between them constitutes the compositional structure of the complex concept. One of the cognitive mechanisms which CMT proposes is employed in concept composition is akin to metaphoric analogy and CMT calls this mechanism *concep-*

tual metaphor. CMT thus takes analogical reasoning to be inherently metaphoric and it takes systematic patterns in linguistic metaphoric expression to be indicative of this cognition. According to CMT, conceptual metaphors can be (1) language-specific, (2) shared among several languages (e.g. within a language family with a shared cultural history), or (3) they can be universal (i.e. occur species-wide around the globe). The idiomatic meanings of metaphoric proverbs are language-specific conceptual metaphors. RT and GS think of inferences as linguistic meanings. On this view, concepts are distinguished from Fregean sense relations. RT and GS therefore use a narrower definition of ‘concept’ than CMT. In Section (2.6), I therefore introduced the notion of linguistic meaning as used by RT and GS. On CMT’s view, idiomatic meanings of metaphoric proverbs are language-specific conceptual metaphors, i.e. they are part of a speaker’s conceptual knowledge. On RT and GS’s view, what makes a linguistic meaning idiomatic is that a particular inference, the idiomatic meaning, is highly salient among other pragmatically plausible inferences and the association between the form of a proverb (its sense relation) and this particular meaning has become conventionalised. Therefore, on this view, idiomatic meanings of metaphoric proverbs are non-conceptual language-specific linguistic conventions, i.e. conventions of conceptualising that are not motivated by or may even go against pure conceptual plausibility, but are motivated by the need to follow linguistic convention in order to facilitate effective communication within a speech community.

According to CMT, inferring metaphoric idiomatic meanings is primarily a matter of conception. According to RT and GS, non-conceptual linguistic conventions play a crucial role in inferring metaphoric idiomatic meanings. The rationale of the experiments in this thesis is that we should choose two languages so that, for the metaphoric proverbs under investigation, the source and target concepts and mental imagery in these metaphors are *cross-linguistically similar* for the two languages. We should try to maximise this cross-linguistic conceptual similarity as much as possible. If CMT is correct, then non-native speakers should be able to infer meanings for these metaphors that are similar to the idiomatic meanings that native speakers infer because of their conception being similar. The more we maximise the cross-linguistic similarity of source and target concepts and mental imagery, the more likely, on CMT’s view, cross-linguistic metaphor communicability and intelligibility should be. If, however, RT and GS are correct, then cross-linguistic metaphor communicability and intelligibility should be *lowered* even when metaphors’ source and target concepts, including their relevance expectations and mental imagery, are cross-linguistically as similar as possible. Speakers can vary greatly in their acquisition of particular concepts because of interpersonal differences in their mental development. In the last part of this chapter (Section 2.7), I therefore discussed to what extent cross-linguistic similarity of source and target concepts is attainable. We wish to compare concepts across languages and since we rely on their representation in natural language (Fregean senses), we need to be aware that (a) a conceptual distinction that is reflected in linguistic expression in one language, might be unmarked in another, and vice versa, (b) an apparent difference in linguistic expression between two languages might not actually indicate a conceptual difference. These two types of non-correlation need to be avoided when selecting metaphoric proverbs with respect to their source and target concepts.

This chapter introduced the major theories concerning the nature of the source and target concepts in metaphors. The next chapter discusses the inferential process of the implied metaphoric analogy.

Chapter 3

Metaphor

3.1 Introduction

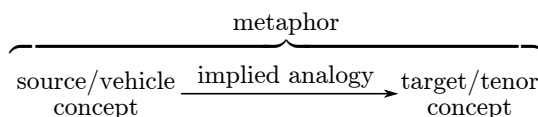
The first part of this chapter discusses inferential metaphor comprehension in general (Section 3.2). Classical rhetoric theory distinguishes between “literal” and “figurative” modes of language use. Grice (1975, 1978) adopted the classical view and proposed that there are distinct inferential processes for the two modes. According to Grice, metaphor comprehension is the result of a failure of direct “literal” attribution of the source concept onto the target concept. Due to a lack of psycholinguistic evidence (Gibbs, 1994a; Keysar et al., 2000; Gibbs, 2002) RT, GS, and CMT reject the literal-figurative distinction as an artefact of linguistic theorising and instead propose that one and the same inferential process underlies comprehension in all uses of language. On their views, to comprehend a metaphor is to infer under which conditions the analogy implied between the source concept and target concept holds. This often means that not all properties of the source concept are transferred to the target concept, but only a subset of them (*semantic narrowing*, Carston, 2002; Wilson & Sperber, 2004) or properties which cannot be said to be part of either the source or the target concept may emerge during comprehension (*emergent properties*, Gineste et al., 2000). The complex meaning of the metaphor that satisfies the conditions of the implied analogy is thus *pragmatically enriched*. A common way of reconciling pragmatic enrichment with the idea that concepts have core meanings is to think of the inferential process as deriving an *ad hoc* concept from the source concept such that its properties match the features of the target concept. It is this *ad hoc* concept and not the original one that is then attributed to the target concept (Barsalou, 1983; McGlone, 1996; Glucksberg, McGlone & Manfredi, 1997; Glucksberg, Manfredi & McGlone, 1997). The inferences necessary to satisfy the metaphoric analogy can vary in inferential strength. Gradient inferential strength is either thought of as *weak to strong implicatures* (Pilkington, 2000) or as *meaning salience* (Peleg & Giora, 2011). These inferences, together with the relevant properties of the source and target concepts (i.e. their *ad hoc* concepts), are a linguistic meaning, specifically *speaker meaning*, the context-dependent meaning implied by the speaker and inferred by other interlocutors.

The second part of this chapter (Section 3.3) discusses how the inferential process changes when metaphors have conventionalised linguistic meanings. CMT assumes that if inferential comprehension is primarily a conceptual process, then when speakers have similar concepts pertaining to metaphors’ language-specific idiomatic meanings (A), they should also comprehend these metaphors similarly (B). CMT raises the stakes of what it means to assume $A \rightarrow B$

because for CMT we are not just making an assumption about the cognitive mechanisms of metaphor comprehension, we are making an assumption about analytical reasoning in general. So while RT, GS, and CMT are interested in understanding the *cognitive* mechanisms underlying inferential metaphor comprehension, CMT risks more by claiming that the mechanisms are not dedicated to language but general mechanisms of analogical reasoning. Ockham’s razor would suggest that if $A \rightarrow B$, CMT offers a simpler, more general, more appealing theory of metaphor comprehension than RT and GS. Interpretations of metaphoric analogies can become conventionalised. Because of the similarity to conventional meanings of idioms, they are also called *idiomatic meanings*. The idiomatic meaning can be thought of as the inferences necessary to satisfy the conditions implied by the metaphoric analogy. RT, GS, and CMT agree that conventionalised idiomatic meanings act as inferential shortcuts: instead of having to consider many potential interpretations and deciding which one is the most plausible in the context of a given metaphor, speakers can directly default to the idiomatic meaning. Idiomatic metaphoric meanings are language-specific. According to CMT, idiomatic metaphoric meanings are sets of language-specific conceptual metaphors. According to RT and GS, they are language-specific conventions of conceptualising that are not motivated by or may even go against pure conceptual plausibility, but are instead motivated by the need to facilitate effective communication. So for CMT, idiomatic metaphoric meanings are conceptual; for RT and GS, they are non-conceptual constraints on conception. We can therefore rephrase the question of whether CMT’s assumption is true by asking whether conceptual or non-conceptual aspects of idiomatic metaphoric meanings are more important in how speakers infer these meanings. The validity of CMT’s assumption that $A \rightarrow B$ depends on the condition that speakers have similar conceptual knowledge, similar concepts, similar mental imagery for the metaphors whose idiomatic meanings we are interested in. The rephrased question depends on the same condition. So testing the validity of the rephrased question is also a test of CMT’s assumption that metaphor comprehension is primarily conceptual. The third part of this chapter (Section 3.4) discusses issues of selecting metaphors with language-specific idiomatic meanings for experimentation.

3.2 The metaphor comprehension process

In cognitive-semantic terms, a metaphor, such as *time is money*, consists of three things: (1) a ‘source’ or ‘vehicle’ concept, MONEY, (2) a ‘target’ or ‘tenor’ concept, TIME, and (3) an ‘implied analogy’ that invites us to comprehend the target concept in terms of the source concept. We can represent a metaphor schematically as follows:



The source concept is obligatory in the linguistic form of a metaphor, but the target concept may be omitted from it. For instance, *black hole* is a metaphor for a type of collapsed star (other types of collapsed stars would, for instance, be *red dwarfs* and *white dwarfs*). *black hole* is the linguistic form that corresponds to the source concept of the metaphor. The target concept is omitted. The full metaphor would be a black hole IS A COLLAPSED STAR WITH A GRAVITATIONAL PULL SO STRONG THAT NO LIGHT CAN ESCAPE IT. If both the source and target concept are represented in the linguistic form, the syntax of a metaphor is

a predicative sentence such as *time is money* or *love is blind*: the objects, the noun (phrase) *money* and the adjective (phrase) *blind*, which correspond to the source concepts MONEY and BLINDNESS respectively, are attributed to the subjects *time* and *love*, which correspond to the target concepts TIME and LOVE. In the grammars of English and German, the predicative syntax of this minimal metaphor form requires the presence of the verb BE. It occurs as either *is* or *are*, in English, and *ist* or *sind*, in German, depending on the number of the subject. Both in the grammars of English and German, this use of BE is called a copula because its only function is to satisfy the syntactic requirement of a sentence to have a finite verb. It is not the auxiliary verb BE as in *I am cooking sprouts for dinner*, but also not the lexical verb BE, as in *there it is*. It can be considered semantically vacuous because it does not have the existential meaning of the lexical verb BE, e.g. *Tom is here*. We can thus say that the syntactic-compositional meaning of a minimal linguistic metaphor consists of two lexical meanings, the lexical meanings of the metaphor source and target concept.

3.2.1 The implied analogy as a failure of direct attribution

On the traditional view of classical rhetoric theory, the syntax of the sentence *time is money*, because it is predicative, suggests that the properties of the concept MONEY should be attributed to TIME in a similar way that in a sentence such as *this house is old*, OLD is conferred onto the HOUSE. However, in the case of *time is money*, the attribution implied by the syntax is not meaningful when taken literally: a house can be old, but to say that time, a physical dimension, can be money, a currency, is, strictly speaking, meaningless. The only way to infer a meaning for *time is money* that is plausible is to situate it in an economic context with the sociocultural background knowledge that payment for work is allotted, among other things, based on the hours of work. In the context of this background knowledge the analogy between TIME and MONEY *is* meaningful. This context is therefore sufficient and necessary to justify inferring the metaphoric meaning, as well as supply the additional information, the information not contained in the mental content of the source and target concepts, to construct this metaphoric meaning, which, because the context is additional information, goes beyond a literal interpretation. This traditional view rests on two assumptions: (1) the default semantic interpretation of the predicative syntax of a sentence such as *this house is old* is one where the object of the sentence, in this case the adjective *old* (and the corresponding concept OLD), is *directly* attributed to the subject of the sentence, *this house* (and the corresponding semantic interpretation that it relates to the concept HOUSE and that in this case, because of the definite determiner *this*, we are talking about a specific instance of the things classed under the concept HOUSE). This direct attribution fails in a metaphor if the source and target concept are taken literal and out of the relevant context. (2) The contextually and pragmatically enriched metaphoric meaning is the result of realising that a literal interpretation is not adequate. It therefore assumes that the linguistic meanings of metaphors are categorically different and distinct from literal meanings. Gricean pragmatics is a the modern version of this traditional view.

Grice follows classical rhetoric theory in that he assumes that linguistic utterances can be literal or figurative (Grice, 1975, 1978). On this view, metaphor belongs to the figurative use of language. The analogy between source and target implied in a metaphor cannot be literally true. On Grice's view, interlocutors usually assume that others in the conversation will do their

best to facilitate effective communication (the Cooperative Principle, in Grice's terms). One way to facilitate effective communication is truthfulness (the conversational maxim of Quality, in Grice's terms). On this view, when a speaker utters a metaphor she deliberately violates, i.e. flouts, interlocutors' expectation of truthfulness. Since the speaker deliberately and obviously flouted truthfulness, other interlocutors seek an interpretation of the speaker's true intended meaning which is in accordance with the source and target concept, a satisfactory pragmatic enrichment of the implied analogy, and which at the same time satisfies their expectation that the speaker meant to cooperate with them through use of the metaphor. Thus, for Grice, the metaphoric meaning intended by the speaker is inferred through an inferential repair strategy that seeks to reconcile that the linguistic utterance at face value is untruthful by pragmatically enriching this initial meaning to the point where it meets interlocutors' expectations according to the maxim of Quality. For Grice, comprehending a metaphor is thus different from comprehending literal utterances and triggers a dedicated inferential process. Relevance Theory (RT) (Sperber & Wilson, 1986; Wilson & Sperber, 1993; Sperber & Wilson, 1995; Wilson & Sperber, 2002, 2004) and Graded Salience (GS) (Giora, 1997, 1998, 1999, 2003; Peleg et al., 2008; Peleg & Giora, 2011) reject the Gricean view that metaphor comprehension employs dedicated inferential mechanisms. They reject the literal-figurative distinction: on their view, the inferential comprehension process is the same for all linguistic utterances; it only so happens that in some cases the inferred speaker meaning differs from the utterance meaning (the meaning initially inferred from the utterance when taken at face value) in just those ways that we would characterise as a metaphor: an attribution (e.g. the property OLD is attributed to an instance of the concept HOUSE) or class inclusion (e.g. in *ice is water*, it is implied that ICE is included in the class of things that the concept WATER describes) is implied but cannot literally be true: for instance, *time is money* implies TIME is a member of the class of MONEY, which is not true when taken literally. Therefore *time is money* is a metaphor, whereas in *ice is water*, ICE is a member of the class WATER, its the solid state of WATER. Hence, the statement *ice is water* is not a metaphor whereas *time is money* is. Hence it is our conceptual knowledge that determines whether we perceive the implied attribution in utterances such as *ice is water* and *time is money* as literal or metaphoric.

I find the 'direct attribution fails' argument problematic for two reasons: (1) it assumes a distinction between 'literal' and 'figurative/metaphoric' linguistic meanings, and (2) it assumes that the default meaning that the predicative syntax of the copula construction implies is direct attribution, i.e. class inclusion. Reason (1), I think, is an artefact of classical rhetoric theory which assumed that there are different modes, different uses of language: a 'literal', 'normal', 'factual' mode and the 'non-literal', 'figurative', 'abnormal' mode, the language of poets. Note the second assumption presupposes the literal-figurative distinction. Classical rhetoric theory divides figures of speech into *tropes* and *schemas*. Tropes are phenomena where the linguistic form is 'normal' but the linguistic meaning is 'abnormal' in the sense that speaker meaning diverges from utterance meaning. Metaphor is a trope. Schemas are phenomena where the linguistic form is 'abnormal' but the linguistic meaning is 'normal'. Alliteration, for instance, is a schema where words are intentionally chosen such that they begin with the same sound. This repetition and rhythmicity is 'abnormal' because it does not normally occur in everyday conversation. This rhythmicity then conveys additional meaning that adds to the otherwise 'normal' meanings of the words chosen; the linguistic meanings of the words chosen are 'normal', i.e. as we would expect in the context of the utterance and conversation and the rhythmicity of

the initial sounds does not change the meanings of the individual words. On this view, schemas have a speaker meaning that contains more information than utterance meaning and the additional information is implied by the formal ‘abnormality’. The *a priori* theoretical assumption of classical rhetoric theory that literal language use is more rudimentary and common is simply not convincing in light of the pervasiveness of figurative and idiomatic language use in natural language. Idiomatic expressions with figurative mental imagery are a frequent feature of natural language use and in idioms the figurative meaning is the most common, most salient linguistic meaning, which most speakers will infer, and most speakers agree that a literal interpretation would be unnatural and semantically implausible. Grice (1975, 1978) held that if inferring figurative meanings requires a dedicated inferential process that differs from the process employed in inferring literal meanings and, at the same time, the inferential process as a whole is such that figurative meanings are inferred only after literal meanings have been inferred and only if the literal meaning does not meet interlocutors’ expectations about speaker meaning, then inferring figurative meanings should require more processing than inferring literal meanings. Gibbs (1994a) found in experiments that processing figurative language is as fast as processing literal language, implying that inferring figurative meanings does not seem to require more processing than inferring literal meanings. Gibbs’ findings thus contradict Grice’s predictions. Contrary to Grice, Gibbs concludes that speaker can infer figurative meanings directly without recourse to literal meanings, in what he calls the direct access hypothesis. Giora (1999) found that Gibbs’ conclusion only holds when the figurative expressions are fairly common whereas if they were unfamiliar novel figurative expressions, it took speakers longer to infer figurative meanings than literal meanings. Giora concludes that figurative meanings can be as inferentially salient as literal meanings and that the driving factor for processing speed is *meaning salience* and not the literal-figurative distinction: salient meanings are processed faster than nonsalient meanings. Salience depends on familiarity, frequency, conventionality, and prototypicality. The more familiar, common, conventional, and prototypical a meaning is, the more likely it is to come to mind more easily and readily, i.e. to be more salient and to be more likely to be inferred (Giora, 1997, 1999, 2003; Peleg et al., 2008; Peleg & Giora, 2011; Giora, Gazal & Goldstein, 2012). Inferential strength, on this view, is then a function of meaning salience. Note that while Grice held that entailments were stronger inferences than implicatures, if figurative meanings are commonly inferred through implicatures and less through entailments, but frequent, common, conventional, and typical figurative meanings can be as salient as literal meanings, then we should conclude that strong implicatures can be inferentially as strong as entailments. The insight that is of interest for the present thesis is that metaphoric meanings, the speaker meaning that interlocutors infer for the analogy between a metaphor’s source and target concept, can vary significantly in inferential strength, depending on how conventional and coded the association between a metaphor’s linguistic form and meaning is: highly conventional metaphoric expressions such as proverbs whose metaphoric meanings are coded as idiomatic meanings have thus highly salient meanings. Novel metaphoric expressions, on the other hand, do not have conventionally coded idiomatic meanings. Rather their meanings are inferred on the basis of speakers’ individual conceptual knowledge and mental lexicons. Hence we should expect a lot of variation in the content of the meaning that different speakers infer for the same novel metaphor. Since speakers cannot rely on conventional idiomatic meanings for novel metaphors, they should be more likely to draw lots of weaker inferences, which, however, together can amount to a conceptually plausible metaphoric meaning. The work of Gibbs and

Giora and their colleagues convinces me to agree with RT and GS that metaphor comprehension requires no dedicated inferential mechanisms but uses the same inferential process as ‘literal’ language use.

Linguistic meaning is highly context-dependent. One might argue that a diagnostic of ‘literal’ meaning is that it prevails when the forms are presented out of the blue, i.e. in the absence of overt context (although this could stem from interlocutors naturally inferring contexts that support this particular meaning rather than others). This diagnostic contrasts with the approach of RT and GS. On the view of RT and GS, one could also argue the ‘literal’ meaning is not context-free, but the meaning that is most frequent across contexts; it is the meaning that is inferred in most contexts. In that sense the ‘literal’ meaning is not context-*free*, it is the most contextual, albeit not context-*dependent*, meaning because there are more contexts in which it is inferred as a salient, plausible, and relevant linguistic meaning. Throughout this thesis I will therefore refrain from characterising one meaning or another of a metaphor’s source and target concept as that concept’s ‘literal’ meaning. Rather I will rank likely inferences in terms of their salience (in GS terms) and contextual relevance (in RT terms).

3.2.2 The implied analogy as indirect attribution via *ad hoc* concepts and pragmatic enrichment

We might call the metaphor concept that emerges from the pragmatic enrichment required in drawing the analogy between source and target concept an *ad hoc* concept if we uphold a literal-figurative distinction. Lawrence Barsalou (1983) introduced the construct of ad hoc categories. *Ad hoc* concepts (Carston, 2002) are cases where concepts are altered or extended temporarily and only within the context of a particular discourse. Consider, for instance, the concept FLUFFY. If concepts have core meanings, what is the core of the concept FLUFFY? It seems that since FLUFFY is a property, which in English is linguistically realised as an adjective, its specific meaning depends on the concept it is applied to, which in English is linguistically realised as a noun. And whenever the nominal concept changes, we derive a new version of the concept FLUFFY and each new version is called an *ad hoc* concept because it is only entertained for as long as it is applied to the nominal concept. For instance, *a fluffy couch* is a different kind of FLUFFY than *a fluffy feather* which again is different from FLUFFY in *a fluffy pillow*. For all we know, there might be an infinite number of concepts that FLUFFY could be applied to as property which would result in an infinite number of *ad hoc* concepts. The derivation of *ad hoc* concepts happens frequently in everyday discourse and as a cognitive task requires much less effort than the internalisation of concepts (e.g. lexicalisation and conventionalisation). Glucksberg, McGlone and Manfredi (McGlone, 1996; Glucksberg, McGlone & Manfredi, 1997; Glucksberg, Manfredi & McGlone, 1997) argue that the implied analogy between the metaphor source and target concept is not achieved through conceptual metaphor, but by what they call *class inclusion*, *attributive categorisation*, or the *attribution model*: from the linguistic expression, which consists at least of the source concept but which may or may not have an overt linguistic form corresponding to the target concept, an *ad hoc* concept is inferred. The source concept exemplifies this *ad hoc* concept. Importantly, they argue that the target concept is attributed to the *ad hoc* concept and not the original source concept. For instance, in *My job is a jail*, the target concept *my job* cannot reasonably be said to belong to the category of BUILDINGS or be an exemplar of the source concept BUILDINGS.

The original source concept JAIL, however, belongs to the category of BUILDINGS. Pragmatic enrichment of the implied analogy, in their view, is then the set of conditions necessary to justify inferring an *ad hoc* concept that can be attributed to the target concept. *Ad hoc* concepts suggest that we need a lot of flexibility in the conceptual system. Recently, CMT has made strong claims about the embodiment of the conceptual system (Lakoff, 2008, 2009, 2013, 2014): (1) all of the physiological evidence suggests that the conceptual system can only be embodied as a fixed neural network in the brain, and that (2) flexibility in the system can thus only be explained through *neural recruitment*, which lends itself particularly well to an explanation of what *ad hoc* concepts are neurologically. CMT can thus handle *ad hoc* concepts through neural recruitment. Categorising at least some metaphor concepts as *ad hoc* concepts is also compatible with Relevance Theory (Carston, 2002). Personally, I am inclined to agree with such a categorisation only if the metaphor concept really is only temporary and highly context-dependent. As such, it seems to me that not all metaphor concepts will classify as *ad hoc* concepts. The proverbs that form the basis for the experiments in this thesis, for instance, do not meet the requirement of *temporariness*. We will see that their interpretation varies with context, but their conventional meanings are certainly not temporary and I would not want to classify their conventional meanings as *ad hoc* concepts. An analysis that involves the notion of *ad hoc* concepts seems more appropriate to the kinds of inferences that arise in comprehending unfamiliar, novel metaphors.

Pragmatic enrichment is more than identifying those properties of metaphors' source and target concepts that are relevant to the analogy implied between them and extrapolating from them in order to construct a metaphoric meaning that meets the conditions of the analogy. In some cases, pragmatic enrichment may involve the addition of conceptual properties. Gineste et al. (2000) claims that during inferential comprehension so called *emergent properties* can arise, properties which are part of a complex compound concept but not of its constituent concepts. For instance, Margolis & Laurence (2011) give the example of a PET FISH which has the feature 'brightly coloured' associated with it. However this feature cannot be said to be necessarily part of either the concept PET or the concept FISH. It must therefore be a property that emerges during inferential comprehension of the compound concept PET FISH. Similarly, Fodor & Lepore (1996) give the example of the compound concept RED HERRING, which is neither 'red' nor 'a herring' as its constituent concepts RED and HERRING imply. Whenever *red* and *herring* appear together in a linguistic utterance, interlocutors who are familiar with its idiomatic meaning must and will not attempt to infer a compositional meaning based on its constituent concepts RED and HERRING but retrieve the idiomatic meaning. Properties that are part of a complex compound concept, such as PET FISH or RED HERRING, can therefore either emerge as part of pragmatically enriching the background assumptions necessary to meet the conditions of the implied metaphoric analogy or they can come from conventionally agreed upon default interpretations (*idiomatic meanings*). These examples show that inferring metaphoric meanings is not just about identifying those properties of the source concept that are likely candidates to be mapped onto matching properties of the target concept (e.g. in the metaphor *time is money*, it is the property that MONEY is a form of work payment and the property that TIME is one aspect of valuing work performance which can be matched up and pragmatically enriched to a point where the analogy implied between them becomes feasible), but also show that in some cases properties themselves emerge during the comprehension process out of a need to meet the conditions of the metaphoric analogy.

According to Relevance Theory (RT) (Sperber & Wilson, 1986; Wilson & Sperber, 1993; Sperber & Wilson, 1995; Wilson & Sperber, 2002, 2004) interlocutors draw those inferences during inferential comprehension that are cognitively most-relevant to understanding the meaning which the speaker is most likely to imply in a given discourse, as well as understanding the speaker's communicative intent in doing so. Assumptions about which information is cognitively relevant to interlocutors is constantly updated as the discourse progresses. Linguistic meaning is thus highly dependent on the context in which it is uttered, the information gathered from the preceding discourse, interlocutors' conceptual and mental lexicons, and the speaker's communicative intent that the interlocutors infer. RT rejects the Gricean view that comprehending metaphors requires dedicated inferential machinery (Grice, 1975, 1978) and instead claims that there is one set of inferential mechanisms that form the machinery to comprehend any and all linguistic utterances. On this view, inferring a metaphoric meaning differs from other linguistic speaker meanings in the salience of those inferences that have to be drawn in order to satisfy the implied analogy. Pilkington (2000) claims that the inferences drawn in metaphor comprehension are more likely to be a large set of weak implicatures rather than few strong inferences (e.g. entailments or strong implicatures) and these weak implicatures accumulate to produce the rich mental imagery associated with the metaphoric analogy in question.

According to Graded Salience (GS) (Giora, 1997, 1998, 1999, 2003; Peleg et al., 2008; Peleg & Giora, 2011) the inferential process consists of two modules: (1) a linguistic module that infers salient meanings based on linguistic knowledge and (2) a pragmatic module that enriches the meaning by taking non-linguistic knowledge such as conceptual, experiential, perceptual, contextual, and world knowledge into consideration. The more familiar, common, conventional, and prototypical a meaning is, the more likely it is to come to mind more easily and readily, and the more salient and likely to be inferred the meaning is. Inferential strength, on this view, is then a function of *meaning salience*. According to GS, the linguistic meaning that is ultimately inferred for an utterance is one that is a compromise between the outputs of the two modules. The two modules are thought to run in parallel so that processing may align at any point during comprehension. This alignment is necessary in order to optimally arrive at a compromise. On this view, metaphor comprehension requires no dedicated cognitive mechanisms beyond the machinery of the two modules and, similar to RT, that an inferred meaning happens to be metaphoric is then a result of the inferences drawn and not the mechanisms involved. Since the second module allows context to be considered, the linguistic meaning that is inferred by the two modules working in unison is *speaker meaning*. GS's view of metaphoric meanings is thus compatible with that of RT.

3.2.3 The implied analogy as a complex concept composed from embodied primary concepts through conceptual metaphors

CMT (Lakoff & Johnson, 1980; Lakoff, 1983, 1987; Lakoff & Johnson, 1999, 2003; Lakoff, 2008, 2009) claims complex inferences, such as linguistic meanings, are the result of conceptual mappings from primary embodied concepts to ever more complex higher-order concepts. Which mappings are established and activated to yield complex inferences, according to CMT, depends on sensory-perceptual stimuli (context, in RT and GS terms) and the mental content and structure of concepts based on past experiences. Although CMT thinks of linguistic meanings as the result of interconnecting primary concepts to ever more complex higher-order concepts,

this view is compatible with RT and GS's view of *speaker meaning* as the context-dependent linguistic meaning implied by the speaker and inferred by listeners. The complex inferences that complex higher-order concepts yield for a particular context, in CMT terms, are the complex inferences that correspond to context-dependent speaker meanings, in RT and GS terms. CMT takes metaphoric-analogical reasoning to be a structural feature of the mental architecture of the conceptual system. Comprehending metaphors, on this view, thus comes natural to the human mind.

CMT argues that implying an analogy in a novel metaphor is achieved through co-activation (synchronised spreading activation) of the source and target concept through neural recruitment to establish a temporary connection (the embodied version of *ad hoc* concepts). By establishing a connection between the two concepts, by virtue of how the central tenet defines complex concepts, the source and target concept together form a more complex concept, the *metaphor concept*. In order to capture all of the variety that can arise from *ad hoc* concepts, emergent properties, and pragmatic enrichment, Conceptual Metaphor Theory must capture this variety (1) in the way that the implied analogy connection is made, (2) which features of source and target it links, and (3) whether the route it takes from source to target incorporates other conceptual mappings along the way. If the temporary connection is used frequently, it becomes permanent (acquisition, learning, and conventionalisation). The neural plasticity necessary to cement new connections diminishes with age. Conceptual Metaphor Theory assumes that most of the mappings are established before the age of five and afterwards learning new conceptual mappings is primarily achieved through neural recruitment and co-activation.

3.2.4 Inferential strength

We can think of idiomatic meanings as linguistic meanings that are strongly implied and inferred. In order to delineate inferences according to their inferential strength it is helpful to distinguish entailments from implicatures. Grice (1975) points out that there are two kinds of inferences: if Elena asks Mike *Are you excited or nervous?* and Mike answers *I'm excited*, then this entails that Mike is either excited or nervous or both. More to the point, it entails that Mike is either excited (and not nervous) or both excited and nervous. It rules out the possibility that Mike is nervous but not excited. Most speakers of English, however, are more likely to understand that Elena can infer that Mike means to say that he is not nervous. Mike could have continued *In fact, I'm nervous as well*, the possibility which entailment rules out. However, notice that Mike could have said *Both*, thus the inference that Mike is not nervous when he answers *I'm excited*, which is an implicature, can be cancelled while entailments cannot be cancelled. Relevance Theory inherits Grice's distinction between *entailment* and *implicature*. Pilkington (2000) makes the important observation that poetry, because it involves complicated metaphors stringed together to elaborate imagery, demonstrates that the inferential process continues for as long as there is a chance of a cognitive effect and in poetic metaphor we not only find implicatures rather than entailments as the primary form of inference, but also implicatures vary in terms of how strongly they are implicated, i.e. in terms of how much we are justified to infer them on the basis of the utterance: strong implicatures are those that are more strongly implicated and that we are more justified to infer and the strength of implication and justified inference then decreases. The important observation that Pilkington makes is that speakers continue processing poetic metaphors even though they often do not strongly implicate, rather

they have many more weak implicatures which together add up to the rich mental imagery that promises a cognitive effect that make inferencing worth continuing.

The distinction between entailment and implicature is important in order to understand CMT's and RT's opposition to an approach to semantics that purely relies on a truth-conditional description of propositions. Entailments, unlike implicatures, are not context-dependent. Approaches to semantics that take linguistic units to only have meaning as truth conditions that are validated based on the external world, exclusively focus on entailments and exclude all inferences that are implicatures. I agree with RT and CMT here that, if we approached semantics in this way, we would have an understanding of *meaning* that was too narrow and we would exclude many, if not the majority of inferences, many of which, as the example conversation between Elena and Mike shows, are much more natural for speakers than the narrow definition of inferences as entailments. RT also often states this position by saying that the object of study should be *speaker meaning*, the meaning that the speaker implicated and intended for us to infer, and not *utterance meaning*, the (literal) meaning that we would attribute to an utterance if we took it at face value and without any knowledge of context or the intentions of the speaker. At the same time relevance theorists would say that speaker meaning is open to a truth-conditional interpretation, at least to some extent. RT's emphasis on the distinction between utterance meaning and speaker meaning is also making the statement that human communication is a process of implying and inferring and not of encoding and decoding: the way we interpret an utterance and the assumptions we make are based on our own experience and we have no direct way of telling what the speaker's original intent was. It is therefore not realistic to assume that speakers can perfectly encode their intended meaning and it is equally unrealistic that listeners could perfectly decode it to infer exactly the meaning intended by the speaker. Speakers' communicative intent then forms an integral part of the meaning communicated, in addition to the content of the utterance itself, because which inferences we draw depends on our interpretation of speakers' intention. Together, I take these insights to mean that it is better to think about inferences as delineated on a scale from strong inferences, the strongest of which we might call entailments, to weak inferences, such as weaker implicatures. The more hypothetical assumptions are that we have to make in order to justify drawing an inference, the weaker that inference is. Alternatively we might say, the more contexts there are that justify drawing a particular inference and the more other speakers draw the same inference, the stronger that inference is. The issue of *justification*, then, becomes a central issue and I agree with Sperber & Wilson (1998) that we need to distinguish between the *mental lexicon* of individual speakers, on the one hand, and the *public lexicon*, the lexicon that corresponds to the strong inferences that most speakers of a language draw in most contexts. We must not take *justification* to mean that there is an institution of speakers who have the authority to dictate what 'standard' inferences of a particular utterance are, but I think we should understand justification in a Lewisian (Lewis, 1969, 1975, 1979, 1983, 1986) and Gricean (Grice, 1975, 1978) sense that the purpose of a communication system is to exchange thoughts and this purpose in and of itself drives speakers to converge on common inferences.

3.2.5 Meaning salience

Meaning salience, the salience of inferences derived during the inferential comprehension process, is greatly influenced by conventional idiomatic meanings which act as default interpreta-

tions. Conceptual Metaphor Theory (CMT) assumes the inferential process takes place primarily on the conceptual level and assumes language-specific conventional default meanings would be incorporated in the mental organisation of those concepts used during their inferential comprehension; Relevance Theory (RT) and Graded Salience (GS), on the other hand, assume that language-specific idiomatic meanings are non-conceptual in nature and their influence on meaning salience shows, so RT and GS assume, that the non-conceptual can overrule the conceptual during inferential processing.

If we reject the traditional distinction between literal and figurative uses of language and instead adopt the position that there is one set of cognitive mechanisms of a general sort that underlie comprehension of all uses of language and at the same time we wish to incorporate the notion that inferences can vary in inferential strength (so that we can characterise idiomatic meanings as strong inferences), it is useful to adopt the notion of salience, the readiness with which some inference comes to mind. Giora & Peleg (Giora, 1997, 1998, 1999, 2003; Peleg et al., 2008; Peleg & Giora, 2011) envision a scale that delineates meanings in terms of *salience*, which they (Giora et al., 2012) define as follows:

A meaning of a stimulus is salient if it is coded in the mental lexicon and enjoys prominence due to factors such as experiential familiarity, frequency, conventionality, or prototypicality; a meaning is less-salient if it is coded but enjoys less prominence due to, for example, reduced exposure or low typicality; a meaning is nonsalient if it is not coded; that is, if it is novel or derivable. Meaning salience, then, is a matter of degree, ranging from coded meanings foremost on our mind to non-coded novel meanings.

(Giora et al., 2012: 24)

Using the notion of *salience*, we can delineate metaphors on a gradient scale from (1) nonsalient: metaphors with very temporary, very context-dependent meanings where the metaphoric meaning is constructed as *ad hoc* concepts, (2) salient: conventional, yet still considered having metaphoric as opposed to literal utterance meanings (e.g. the metaphoric proverbs that will be a central focus in this thesis) to (3) highly salient: those metaphors that are traditionally called *dead metaphors*, i.e. metaphors that have been conventionalised to such a degree that speakers consider the metaphoric meaning *to be* the literal utterance meaning. Note that this scale presupposes a distinction between *literal* and *metaphoric* meaning. Recall that in Conceptual Metaphor Theory mappings between concepts are made through *similarity matching* which is inherently like metaphoric analogy. Conceptual Metaphor Theory therefore rejects the classical distinction between literal and metaphoric meaning and replaces it with the proposal that much of the conceptual circuitry is metaphoric-analogical via *similarity matching*. CMT's view should therefore be compatible with GS's view of delineating inferences in terms of their salience. In CMT's terms, conceptual metaphors correspond to salient metaphoric meanings.

If inferential strength correlates with salience, then the more salient an inference is, the inferentially stronger that inference is. If what determines the magnitude of inferential strength is the amount of justification for drawing it, then the amount of justification should also correlate with salience: an inference is more salient the more justified we are in inferring it. For any given inference we should then be able to ask, in the given context, what is salient about the inference that justifies drawing it, that justifies expending cognitive effort? In RT terms, what

we are asking is: what is the positive cognitive effect that inferring it would have that warrants spending cognitive effort on inferring it? For a metaphor, which minimally consists of a source indirectly attributed to a target concept, ‘salience of the inference’ should then be taken to be the salience of those features of the source concept, i.e. the features of the derived *ad hoc* concept, that can be attributed to the matching (salient) features of the target concept. Salience of a metaphor should thus be the salience of the implied analogy, i.e. the salience of those features of the *ad hoc* source concept that we are justified to apply to the target concept. But of course we must not forget that salience here (Giora, 1997, 1998, 1999, 2003; Peleg et al., 2008; Peleg & Giora, 2011) is not just conceptual, but also incorporates familiarity and conventionality as a result of the frequency of linguistic use and the prototypicality of concepts’ features. Prototype theory seems like a natural candidate for a theory to predict which features of concepts should be more salient than other features. Prototypical-salient features should be more likely to be transferred from *ad hoc* source concepts to target concepts.

RT captures the idea of inferential strength in the cognitive notion of *relevance*: an inference is the stronger, the more cognitively relevant it is in the current discourse to bridge the disjunct of utterance and speaker meaning through pragmatic enrichment. RT defines cognitive relevance in terms of the notions of *cognitive effort* and *cognitive effects*: cognitive effort is the amount of effort expended during inferential comprehension. The smaller the cognitive effort required to draw inferences, the greater their cognitive relevance is. On this view, the inferential process applies cognitive comprehension mechanisms in order to draw inferences that will increase the cognitive relevance of the speaker meaning that is the result of the process. Cognitive effects are the cognitive benefits that drawing particular inferences have toward that goal. Cognitive effects can be positive or negative. Positive cognitive effects are the result of drawing inferences that increase the overall relevance of the speaker meaning thus-far; negative cognitive effects are the result of drawing inferences that decrease the overall relevance of the speaker meaning inferred. The greater the cognitive effects that are achieved by processing and input to the inferential process, the greater its relevance is (Sperber & Wilson, 1986; Wilson & Sperber, 1993; Sperber & Wilson, 1995; Wilson & Sperber, 2002, 2004).

Linguistic meaning is well-known to be highly context-dependent. Upon frequent exposure to conventional linguistic expressions that result in drawing the same inferences over and over again, these inferences increase in meaning salience, inferential strength, cognitive relevance and, in RT terms, lead speakers to form expectations of contextual relevance, i.e. speakers form expectations as to which inferences they associate with which inferential contexts. We can therefore assume that conventionalisation of metaphors’ linguistic meanings to idiomatic meanings also leads speakers to form expectations as to which contexts they commonly associate with these metaphors. CMT takes a similar view on this: metaphoric idiomatic meanings are complex concepts that correspond to the set of inferences drawn in during inferential comprehension and this complex concept is derived from primary embodied concepts through conceptual metaphors which, because the meaning is conventional, are also conventionally established and accessed. The set of primary embodied concepts is CMT’s equivalent notion of RT and GS’ notion of context. Primary embodied concepts are activated upon perceptual experience. Depending on the context of the discourse, only some of the primary embodied concepts associated with the complex concepts are activated. The metaphoric meaning is thus modulated according to context. The firmness of the neural mappings and gating with which certain primary embodied concepts and particular conceptual metaphors are interconnected

to form the complex concept of the idiomatic meaning determines the salience and strength of inferences and expectations of contextual relevance. As such, CMT, on the one hand, and RT and GS, on the other, should make the same experimental predictions with regards to the effects of meaning salience, inferential strength, context, and relevance expectations.

How cognitive effort, effects, relevance, and salience can be quantified is a controversial issue in linguistic theorising, but all that is required for this thesis, is that we can assume that they can quite often be compared for inferences. RT supports this functional operationalisation of inferential comparability:

[R]elevance theory does not provide an absolute measure of mental effort or cognitive effect, and it does not assume that such a measure is available to the spontaneous workings of the mind. What it does assume is that the actual or expected relevance of two inputs can quite often be compared. These possibilities of comparison help individuals to allocate their cognitive resources, and communicators to predict and influence the cognitive processes of others. They also make it possible for researchers to manipulate the effect and effort factors in experimental situations.

(Wilson & Sperber, 2004: 39–40)

While I have tried to emphasise the similarities between RT and GS, on the one hand, and CMT, on the other, it is important to remember that conventionalisation in RT and GS is a *cultural process* while temporary neural mappings becoming permanent is a *neural process* in CMT. So with respect to these processes, while linguistic accounts of learning and neural accounts of learning should come to an agreement because they both focus on the individual speaker, conventionalisation includes the social and cultural dynamics which we can think of as *additional* pressures on the individual learner to acquire and cement particular mappings. In this thesis we will see that the cultural component has a tremendous effect on how speakers structure concepts and conceptual mappings (*idiomatic* metaphoric meanings) and how they facilitate them when processing metaphors that are *unfamiliar* to them.

3.3 Metaphors with conventional meanings

Using the notion of salience we can delineate metaphoric meanings from nonsalient to salient: (1) nonsalient metaphors have very temporary, very context-dependent meanings where the metaphoric meaning is constructed as *ad hoc* concepts, (2) salient metaphoric meanings are conventional and very likely to be inferred. We can subdivide salient metaphoric meanings into (2a) those that are still considered ‘metaphoric’ as opposed to literal utterance meanings, and (2b) the meanings of those metaphors that are traditionally called *dead metaphors*, i.e. metaphors that have been conventionalised to such a degree that speakers consider the metaphoric meaning *to be* the literal utterance meaning. We can also delineate metaphoric meanings in terms of their conventionality: (1) novel metaphors have unconventional meanings and the inferential process is likely to be focussed on utterance meaning and speakers’ individual mental content and representation of source and target concepts. (2) Conventional metaphoric meanings are linguistic meanings where the context that would be necessary to infer this meaning as an utterance meaning in a non-idiomatic way is encapsulated in the conventional meaning. Similar to the graded salience scale, we may subdivide conventional metaphoric

meanings into (2a) those still considered ‘metaphoric’ and literal utterance meanings are available as inferential alternatives, and (2b) the meanings of dead metaphors where literal utterance meanings are not available as inferential alternatives.

CMT proposes that idiomatic meanings such as those of metaphoric proverbs are conceptual metaphors, i.e. the metaphoric meanings are complex concepts that are inferred through conceptual metaphors that derive the complex concept from primary embodied concepts. CMT claims that analogical reasoning is thus inherently metaphoric and consequently CMT sees metaphor as a cognitive phenomenon and not as a linguistic one. However, RT and GS see metaphor as a primarily linguistic phenomenon. RT and GS, like CMT, recognise that the precise mental content and representation of metaphors’ source and target concepts and the structure of associations within the mental lexicons of individual speakers is important in order to understand metaphor comprehension, but contrary to CMT, RT and GS emphasise that idiomatic meanings of metaphors such as metaphoric proverbs involve linguistic conventions and these conventions are non-conceptual, i.e. not reducible to the cognitive processes of conceptualising. If, as CMT proposes, cross-linguistic approximate conceptual dissimilarity is more likely between two languages with little or no linguistic contact and little to no common cultural background, then the reverse should also be true: cross-linguistic approximate conceptual similarity should be very likely between two languages with a lot of linguistic contact and a rich shared cultural background. If, as CMT argues, metaphor comprehension relies on cognitive mechanisms of analogical reasoning that are inherently like metaphors (conceptual metaphors) and metaphoric meanings are complex concepts, then cross-linguistic metaphor intelligibility should be more likely between languages with approximate conceptual similarities because of sociocultural interaction, history, and exchange of ideas (concepts). In order to test the competing claims of CMT, on the one hand, and RT and GS, on the other, regarding the nature of metaphoric idiomatic meanings, we need something to compare the cross-linguistic intelligibility of the metaphors in question with. If idiomatic meanings require non-conceptual knowledge of linguistic conventions, as RT and GS claim, then we need to compare cases where this knowledge is available to speakers with cases where that knowledge is not available.

There are no conventional idiomatic meanings available for novel metaphors. Hence, when speakers encounter these unfamiliar novel metaphors, they must rely entirely on their judgements of conceptual plausibility while they draw the inferences necessary to facilitate *ad hoc* concept derivation and pragmatically enrich the implied analogy. Importantly, linguistic convention can only influence their judgements on the lexical level of metaphors’ source and target concepts in this case, but not on the level of the implied analogy because there is no established idiomatic meaning for these novel metaphors.

Novel metaphors make for an excellent control condition when investigating the nature of idiomatic meanings of language-specific metaphoric proverbs: CMT proposes that complex concepts, which correspond to the idiomatic meanings, are derived from primary embodied concepts through conceptual metaphors. These conceptual metaphors establish the necessary conceptual modification of the primary concepts and intertwine them to more complex concepts, which then, on CMT’s view, correspond to complex linguistic inferences. Which modification are made and which interconnections are established, CMT claims, is guided by a cognitive mechanism akin to metaphoric similarity matching. This cognitive mechanism of metaphoric conception, CMT argues, is more fundamental than and independent of non-conceptual linguistic conventions which are specifically linguistic and might enter the inferential comprehension

process at some point. Contrary to CMT, RT and GS propose that metaphor comprehension is inherently linguistic and might involve linguistic non-conceptual conventions where these conventions are constitutive of the idiomatic meaning and therefore linguistic convention plays an inherent role in inferential metaphor comprehension, a role that supersedes pure conception.

If CMT is correct and metaphor comprehension is primarily a matter of conception and linguistic conventions are only secondary, then idiomatic meanings should conceptually relate to their constituent source and target concepts. On this view, even speakers who are unfamiliar with the linguistic conventions that give rise to the idiomatic meaning should nevertheless be able to infer a non-idiomatic meaning approximately conceptually similar to the true idiomatic meaning on the basis of source, target, and their general conceptual knowledge alone. If RT and GS are correct and comprehension of metaphors with idiomatic meanings is primarily a matter of speakers' awareness of linguistic conventions pertaining to the idiom and only secondarily a matter of conception, then the difference in mental content between the non-idiomatic meaning that speakers who are unfamiliar with the idiomatic meaning infer and the true idiomatic meaning should be greater than we would expect if CMT was correct.

In the case of novel metaphors there are no conventionally established idiomatic meanings associated with them yet. There might be idioms that speakers might consider during inferencing because they are relevant to the semantic fields that the source and target concepts of the novel metaphors activate, but otherwise comprehending novel metaphors should consist of retrieving the lexical concepts of the source and target, derivation of *ad hoc* concepts from the source to facilitate indirect attribution to the target, and pragmatically enriching the mental content of the *ad hoc* concept in order to satisfy the conditions of the implied metaphoric analogy as defined by expectations of contextual relevance. Since novel metaphors have no conventionally established idiomatic meanings, all speakers whose mental lexicons are approximately conceptually similar for the source and target concepts in question, who share approximately similar world knowledge, and who come from a similar cultural background should thus be likely to draw approximately similar inferences and the non-idiomatic meanings they infer for the novel metaphors are thus more likely to be more approximately conceptually similar than if they were given a random sample of metaphors along the graded salience scale. Therefore, if RT and GS are correct and the non-idiomatic meanings that speakers who are unfamiliar with the idiomatic meanings of language-specific metaphoric proverbs infer are different from proverbs' true idiomatic meanings, then the nature of the inferential process of the unfamiliar speakers should be similar to comprehending novel metaphors. If CMT is correct and this difference in inferential content between idiomatic and non-idiomatic meanings should be small or non-existent (in the sense of approximate conceptual similarity) for language-specific metaphoric proverbs, then the nature of inferential comprehension of novel metaphors should not differ from comprehending language-specific metaphoric proverbs. Novel metaphors are therefore a theory-neutral control condition in the experiments of this thesis.

3.3.1 Idiomatic meanings as inferential shortcuts

The idiomatic meaning of a metaphoric proverb can be understood as an inferential shortcut: instead of having to consider all conceptually possible alternative interpretations and assessing their plausibility in the context of the analogy implied by the metaphor, speakers who are familiar with the idiomatic meaning are provided with a default interpretation by it, which

eliminates the need to consider all alternatives. The idiomatic meaning therefore is a salient and preferred meaning.

Processing effort and time should decrease whenever inferential shortcuts are used. Keysar et al. (2000) assume that when particular conceptual mappings are required in order to comprehend a linguistic expression, then when these mappings are not available to speakers, they should take longer to comprehend. Keysar et al. (2000) found no differences in processing speed between expressions for which CMT proposes that speakers must rely on conceptual metaphors and novel expressions where no conceptual metaphors should be available as inferential shortcuts. They therefore conclude that conceptual mappings of the kind that CMT proposes might not exist. Contrary to Keysar et al. (2000), speakers show slower reading/response times for novel metaphors with unconventional linguistic meanings, and faster reading/response times for metaphoric proverbs with familiar idiomatic meanings (see Chapter 7, Section 7.4). I therefore think it is reasonable to assume that processing time correlates with the use of inferential shortcuts. Giora (1999, 2003); Peleg et al. (2004, 2008); Peleg & Giora (2011) showed that metaphors with salient conventional meanings are processed faster than when meanings are nonsalient and unconventional. Therefore, processing speed seems to be correlated with salience. Salient default interpretations such as the idiomatic meanings of the metaphoric proverbs under investigation in this thesis should thus be comprehended faster than novel metaphors without conventional meanings.

The key point here is that the inferential process of novel metaphors might be quite different from that of metaphors with conventional idiomatic meanings in terms of the cognitive mechanisms involved. For novel metaphors, which have no conventional meanings associated with them, speakers have to explore all possible interpretations by exploring multiple ways (inferential routes) to enrich the meaning of the source and target concept (thereby deriving *ad hoc* concepts) through a wide range of weak implicatures that flesh out a potential set of conditions that can satisfy the implied analogy. On the other hand, when conventional idiomatic meanings are available, the conventional meaning supplies a set of implicatures that amount to a default mental image and contextual relevance expectations, and thereby eliminates the need to infer other implicatures. Inferring other implicatures for alternative interpretations would be wasting cognitive resources because these implicatures would be weaker than those associated with the default interpretation. Because of the salience of the idiomatic meaning, the implicatures associated with it are also inferentially stronger. The metaphoric proverbs in the experiments of this thesis have idiomatic meanings that are either conventionalised in both English and German or in only one of the two languages. The novel metaphors that only resemble the linguistic form of the proverbs, however, do not have any conventionally established idiomatic meanings. With respect to those proverbs whose idiomatic meanings are specific to either English or German, we can therefore test whether non-native speakers will process them as if they were novel metaphors or as if they were conventional metaphors.

3.3.2 A general theory of inferential comprehension

In classical rhetoric theory (Aristotle), literal language use is seen as the default; figurative language use is seen as an “abnormal” and “unusual” use of “literal” language. Classical rhetoric theory distinguished two types of “abnormal” use (1) “normal” lexical and compositional meaning but “abnormal” linguistic form (*schemas*) and (2) “normal” linguistic form but “abnormal”

meaning (*tropes*). Metaphor was categorised as a trope. Grice (1975, 1978) continued this tradition. He characterised metaphor as “flouting” the maxim of Quality, i.e. metaphor is characterised as a deliberate lie because, when taken at its “literal” face value, the implied metaphoric analogy is never true. The first step in comprehending a metaphor, on Grice’s view, is to understand that an analogy is not literally given but only implied, implied because we must assume that the speaker is also aware that direct literal attribution fails and would not have uttered the metaphor if the literal meaning was the meaning intended by the speaker. On the relevance-theoretic view, the first step in inferential comprehension is to recognise the speaker’s communicative intent. Grice, contrary to RT, assumes that first listeners infer the literal meaning and only upon realising that it is at odds with the context in which it was uttered, recognise the speaker’s communicative intent and then go about inferring a pragmatically enriched interpretation of the implied metaphoric analogy, a meaning that is in agreement with the speaker’s intent and context. Due to lack of psycholinguistic evidence the literal-figurative distinction has been abandoned as a theoretical construct (Gibbs, 1994b,a, 1996, 1998; Gibbs & Steen, 1999; Keysar et al., 2000; Colston & Gibbs, 2002; Gibbs, 2002). RT rejects the idea that any linguistic meaning can be literal, but argues that all meaning is inherently contextual and what the classical view calls a “literal” meaning is just a meaning that occurs (a) in most contexts and (b) on which most speakers of a language community agree and converge during inferencing out of convention. GS introduces a model of meanings varying in terms of their conventionality. Similar to RT, on GS’ view, a “literal” meaning is a meaning that depends least on context and primarily on conventionality (salience).

There is a similar reversal in perspective: in traditional semantics, language is taken to use abstract symbols and abstract meanings which may or may not refer to the actual world. Abstract concepts are assumed to be the default; concreteness is relegated to referential semantics, the relation between these abstract symbols and reality. Embodied Cognition (Gibbs, 2003, 2006), CMT in particular, assume concrete perceptual concepts are the default (called *primary embodied concepts* in CMT); abstract concepts are created by combining concrete concepts. The more concrete concepts have to be combined, the more abstract the complex concept appears to us, CMT argues. The reversal in the perspective on abstract and concrete concepts is a reversal in the theory of conception. CMT claims that the way that concrete basic concepts are combined into more complex, potentially abstract, concepts is through conceptual metaphors (conceptual mappings reminiscent of metaphoric analogy based on similarity matching). On this view, much of cognition is inherently metaphoric. Linguistic metaphor is not seen as a distinctly linguistic phenomenon because it is taken to simply use the cognitive mechanisms of conceptual metaphor. “Literal” language use often involves complex and abstract concepts. On CMT’s view, “literal” language use is therefore seen as an elaborate case of “figurative” language use. So CMT, in a lot of ways, takes figurative language use and metaphor to be more “normal” than “literal” language use. CMT thus takes the opposite view on modes of language use to classical rhetoric theory and the opposite view on abstract and concrete concepts; two reversals in perspective. Analogical reasoning is one of the central types of human reasoning. For CMT, investigation of metaphor is an investigation of the cognitive mechanisms underlying analogical reasoning because for CMT metaphor not just as a phenomenon of rhetoric but as a cognitive mechanism that is characteristic of analogical reasoning.

Following Grice’s introduction of the entailment/implicature distinction, pragmatics emerged as a new field of study, with an object of study distinct from that of semantics. Semantics deals

with entailments and generalisations of their inference; pragmatics deals with implicatures and their inferential generalisations. Entailments are inferences that are strictly necessary for an utterance to be true. An implicature is that which is suggested by an utterance but not a condition for its truth. The split between semantics and pragmatics assumes that the inferential generalisations for entailments and for implicatures are different. CMT, RT, and GS agree that entailments are the strongest inferences, but emphasise that they can be delineated on one scale with implicatures according to their inferential strength (meaning salience). According to Grice, the inferential process has two separate parts: one process that infers literal meanings using only entailments and another process that infers contextual meanings using implicature. Since Grice assumes that the two processes are completely separate, he also assumes that there are different sets of rules guiding each, different generalisations. We thus get a split between the domain of semantics and pragmatics. Semantics is about the inferential rules of literal meanings, which involve entailments and strict conditions of logical and factual truth. Pragmatics is about the inferential rules of meaning in context, which involve implicatures and pragmatic rules, e.g. the Gricean maxims (Grice, 1975, 1978) and Searle's inferential rules of speech acts (Searle, 1969, 1975, 1979). To assume that there are different sets of inferential rules for entailments and implicatures would presume that the inferential process would have to start with the decision which of the two sets should be used, which, however, would require us to know whether the result should be an entailment or an implicature (a chicken and egg problem of infinite regress). This is only a problem if there are clashes between the outputs of the rules. We could alternatively assume that it is quite possible for interlocutors to have deductive inference producing semantic consequences and alternative-based reasoning producing pragmatic consequences in operation at the same time. GS comes closest to this view when it proposes that two inferential modules, one concerned with semantic consequences and the other concerned with pragmatic consequences, work in parallel and unison to arrive at a compromise. Grice claims that the inferential process always infers using the semantic rules of entailments first and only when this result conflicts with the discourse context, is a contextual meaning inferred using the rules of implicatures. Experiments have found no difference in processing speed between inferences that should only require the rules of entailments and inferences that should also involve the rules of implicatures (Gibbs, 1994a; Keysar et al., 2000; Gibbs, 2002). CMT, RT, and GS reject the idea that there are two parts to the inferential process that work according to different sets of rules. They therefore reject that generalisations about entailments are *fundamentally* different from generalisations about implicatures. Instead they postulate *one* set of rules that are supposed to guide *all* of the inferential process. They therefore must also reject the idea that semanticists and pragmaticists can study inferencing without comparing notes. So although their methodologies may seem pragmatic, it would be wrong to characterise their theories of inferential processing as pragmatic theories. We might say their theories are on the semantics-pragmatics interface, but it is important to realise that they are intended as holistic theories of inferencing that cover both entailments and implicatures with one set of inferential rules. CMT, RT, and GS thus incorporate both the domains of semantics and pragmatics. CMT, RT, and GS use the adjective "cognitive" to signify that the object of study is this holistic set of inferential rules. So calling the rules cognitive is a way to distinguish them from semantic-inferential rules that only cover entailments and pragmatic-inferential rules that only cover implicatures.

A theory of inferential comprehension that treats concepts as atomic symbols which are

only recombined compositionally in a propositional language of thought is not appropriate for metaphor comprehension because metaphor comprehension involves concept modulation: not all features of the source concept are implied to map to all features of the target concept. Therefore, the objective must be to find generalisations of *conceptual modulation*. RT uses the terms *semantic narrowing* and *broadening* to talk about this modulation. GS talks about it in terms of meaning salience affecting the interaction between the two inferential modules. CMT is an entire theory built around the idea of conceptual modulation. CMT, RT, and GS recognise the tension between inference guided by conventionality and non-conventional, free contextual inference and they therefore make meaning salience a central focus. They also recognise conceptual modulation. They aim to postulate theories of inferential processing that can account for conceptual modulation, even make generalisations about it, *in light of* the tension between context and convention in meaning salience. The contextuality of inferential meaning must play a central role in a theory of inferential comprehension. In such a theory, conventionality takes the form of inferential shortcuts, shortcuts which eliminate the need to consider many contextually plausible interpretations or at least reduce the number of options considerably. The central question of a theory of inferential comprehension then becomes “How much (contextual and conventional) information do you need to consider before you can infer a plausible meaning?” Meaning salience then follows as a function of the amount of the information considered. The inferential process is parsimonious and aims to infer salient meanings with the least amount of cognitive effort. Considerations of cognitive effort provide a natural stopping point: the inferential process continues until the cognitive effort required to continue the process exceeds potential gains from considering more information. Meanings inferred through such a process are open to truth-conditional interpretation, but as inferential strength decreases, conditions of truth become less strict. CMT, RT, and GS envision theories of inferential comprehension that acknowledge the importance of all of these characteristics.

The semiotic triangle of thought, utterance, and the world (de Saussure, 1993, 1916) allows for different definitions of semantics. Semantics may be defined as the relation between an utterance and the world. On this view, truth conditions are conditions of reference: any proposition may be uttered, but only those that adequately refer to reality are true. This approach to semantics is also called propositional, referential, or representational semantics. Semantics can also be defined as *narrow semantics*, a term introduced by Micheal McDermott (1988). Narrow semantics is the relation between utterance and thought (Fodor, 1975; Lakoff & Johnson, 1980; Chomsky, 1996; Fodor, 2008; Lakoff, 2009). On this view, truth conditions are conditions of inference: many thoughts may be entertained, but only when they are inferable from an utterance is its meaning true. The relation between thought and the world is then what is inferred during the comprehension process. CMT, RT, and GS agree that a referential-semantic approach to metaphor comprehension is the wrong approach, not because it could not describe metaphoric propositions, but because in doing so we would not arrive at any generalisations. Propositional-referential semantic theories are unsuitable for explaining metaphor comprehension because they ignore conceptual modulation: each modulation is assigned a distinct propositional representation and distinct mapping to a referent, instead of revealing which modulations are related to each other. However, revealing which modulations are related to each other would precisely be the kinds of general inferential rules of metaphor comprehension we are looking for. Also, referential semantics offers no tools to distinguish between conceptual and non-conceptual aspects of idiomatic meanings, the central issue of this thesis.

If metaphor comprehension is about the modulation of conceptual knowledge, does that mean that conventionality in metaphor is also conceptual? CMT thinks the answer is “yes” because it takes semantic roles (e.g. SOURCE, PATH, GOAL, INTERIOR, EXTERIOR) not just as labels of semantic functions within a theory but computational building blocks that can be combined compositionally, in connection with conceptual metaphors (e.g. ABSTRACT IS CONCRETE), to yield inferences. RT and GS think the answer is “no” and point to the possibility that speakers’ convergence on conventional meanings may simply be them following the cooperative need to facilitate communication. CMT assumes that if inferential comprehension is primarily a conceptual process, then when speakers have similar concepts pertaining to metaphors’ language-specific idiomatic meanings (A), they should also comprehend these metaphors similarly (B). CMT raises the stakes of what it means to assume $A \rightarrow B$ because for CMT we are not just making an assumption about the cognitive mechanisms of metaphor comprehension, we are making an assumption about analytical reasoning in general. So while RT, GS, and CMT are interested in understanding the cognitive mechanisms underlying inferential metaphor comprehension, CMT risks more by claiming that the mechanisms are not dedicated to language but general mechanisms of analogical reasoning. Ockham’s razor would suggest that if $A \rightarrow B$, CMT offers a simpler, more general, more appealing theory of metaphor comprehension than RT and GS. If $A \rightarrow B$, then the opposite, $\neg B \rightarrow \neg A$, should also be true (the law of contraposition): $(A \rightarrow B) \iff (\neg B \rightarrow \neg A)$. Conventionalised idiomatic meanings act as inferential shortcuts: instead of having to consider many potential interpretations and deciding which one is the most plausible in the context of a given metaphor, speakers can directly default to the idiomatic meaning. Idiomatic metaphoric meanings are language-specific. According to CMT, idiomatic metaphoric meanings are sets of language-specific conceptual metaphors. According to RT and GS, they are language-specific conventions of conceptualising that are not motivated by or may even go against pure conceptual plausibility, but are instead motivated by the need to facilitate effective communication. So for CMT, idiomatic metaphoric meanings are conceptual; for RT and GS, they are non-conceptual constraints on conception. We can therefore rephrase the question of whether CMT’s assumption is true by asking whether conceptual or non-conceptual aspects of idiomatic metaphoric meanings are more important in how speakers infer these meanings. The validity of CMT’s assumption that $A \rightarrow B$ depends on the condition that speakers have similar conceptual knowledge, similar concepts, similar mental imagery for the metaphors whose idiomatic meanings we are interested in. The rephrased question depends on the same condition. So testing the validity of the rephrased question is also a test of CMT’s assumption that metaphor comprehension is primarily conceptual.

3.3.3 Toward testing RT, GS, and CMT’s predictions concerning metaphoric idiomatic meanings

CMT claims that idiomatic meanings are conceptual metaphors. Some conceptual metaphors are universal, i.e. they occur species-wide in all languages. Other conceptual metaphors occur only in particular languages. A common cultural history or cross-cultural linguistic contact makes it more likely that language-specific conceptual metaphors are culturally transmitted from one language to another. If by ‘idiomatic’ we mean that these meanings of metaphoric proverbs are language-specific, they should not be universal but language-specific conceptual metaphors according to CMT. RT and GS also claim that idiomatic meanings of metaphoric

proverbs are language-specific, but language-specific linguistic conventions, conventions that are specifically non-conceptual, i.e. they are not motivated by how speakers conceptualise the world around them but by the need to facilitate effective communication. On this view, speakers acquire, internalise, and expect other interlocutors to be familiar with the idiomatic meaning because the idiomatic meaning is an inferential shortcut. What speakers who use the metaphoric proverbs expect from other interlocutors is that they will retrieve the idiomatic meaning from their mental lexicons *without* actually evaluating the conceptual plausibility of the implied analogy and pragmatically enriching that analogy to a point where they feel contextually justified making the necessary additional assumptions. The other interlocutors do not really need to evaluate whether the analogy is conceptually plausible and whether it warrants making those assumptions as long as they know the idiomatic meanings through linguistic convention. The other interlocutors are given all of the necessary assumptions, the fully enriched meaning as part of the idiomatic meaning. If RT and GS' view of idiomatic metaphoric meanings is correct, then speakers might actually overestimate the conceptual plausibility of idiomatic meanings: they might think and have strong intuitions that the idiomatic meaning of a proverb is conceptually plausible when in fact it is linguistic conventions that gives them that impression. It might be that speakers are unable (or at least find it quite hard) to introspectively differentiate between (a) their mental content and representation of the source and target concept and the conceptual plausibility of the analogy based on their knowledge of the features of the source and target concept alone, and (b) the additional assumptions that are added to (a) in order to yield the full meaning of the metaphor. On CMT's view, when the idiomatic meaning contains more information than the sum of the source and target concept, then that additional information must also be supplied via the metaphor circuitry that is part of the conceptual metaphor that is the neurocognitive correlate of the idiomatic meaning. But note that CMT would argue that the additional information would be part of speakers' conceptual knowledge whereas RT and GS would call the additional information 'non-conceptual'. CMT would also expect, like RT and GS, that speakers might be unable to differentiate between the additional information and the rest of the metaphor circuitry that relates to a metaphor's idiomatic meaning. In both theoretical frameworks that I wish to experimentally test in this thesis, CMT, on the one hand, and GS and RT, on the other, we thus face the possibility that speakers are unable to introspectively differentiate between (a) their knowledge of the source and target concept and the conceptual plausibility of the analogy and (b) the additional information necessary to infer the idiomatic meaning.

What we thus need is a way to disentangle (a) and (b): We want to compare how two groups of speakers process a set of metaphoric proverbs with idiomatic meanings: speakers in both groups should have (a), i.e. have approximately conceptually similar mental representations for source and target concepts and should be likely to select similar conceptually salient features of those concepts to derive *ad hoc* concepts and indirectly attribute them to the target concept, but only speakers in one of the groups should be familiar with (b). One way to ensure that only speakers in one group will know (b) should be if only these speakers are familiar with the idiomatic meanings. If, as CMT claims, (b) should be part of speakers' conceptual knowledge and their mental lexicons, then speakers in the other group who do not know the idiomatic meanings of the respective metaphoric proverbs should nevertheless be able to infer linguistic meanings that are approximately conceptually similar to the idiomatic meanings by virtue of their knowledge of (a). Since CMT claims that idiomatic meanings are language-specific

conceptual metaphors, it should thus be more likely that speakers of both groups infer similar meanings if they are speakers of two languages with a shared cultural history and close exchange of concepts through cultural transmission and linguistic contact. If CMT is correct in this prediction, then if we chose to compare two languages in close linguistic and cultural contact and whose speakers cross-linguistically have approximately conceptually similar source and target concepts for the metaphoric proverbs in question, then non-native speakers should be able to infer meanings similar to the idiomatic meanings that native speakers know. If RT and GS are correct in their predictions, even in this situation where we maximise the chance for cross-linguistic metaphor intelligibility by maximising the cross-linguistic approximate conceptual similarity of the mental lexicons with respect to (a) for speakers of the two languages and where we have ensured closeness of cultural and linguistic contact, then we should nevertheless be able to find language-specific metaphoric proverbs for which non-native speakers infer meanings that are unlike the idiomatic meanings of native speakers. The experiments in this thesis thus aim to maximise cross-linguistic approximate conceptual similarity and look for lowered cross-linguistic metaphor intelligibility. Cross-linguistic metaphor intelligibility will be lower if the meanings that non-native speakers infer are different from the idiomatic meanings. Cross-linguistic metaphor intelligibility will also be lower if knowledge of (b) leads native speakers to overestimate the conceptual plausibility of idiomatic meanings: if so, then the conceptual plausibility of language-specific metaphoric proverbs should be lower when judged by non-native speakers than when judged by native speakers, even when knowledge of (a) and cross-linguistic conceptual, cultural and linguistic closeness are maximised.

In this thesis we are interested in the cross-linguistic intelligibility of language-specific metaphoric proverbs. Proverbs have salient idiomatic meanings as strong conventional default inferences. I proposed that we use novel metaphors as a control condition. There are no established conventional idiomatic meanings for novel metaphors yet. Hence speakers must fully rely on context and their conceptual knowledge regarding the source and target concepts. If RT and GS are correct and idiomatic meanings require speakers to know non-conceptual linguistic conventions, then when this knowledge is not available to speakers their inferential comprehension process should operate in the same way it would for novel metaphors. Metaphoric proverbs have conventional meanings as strong inferences. If we seek to compare the comprehension process of conventional metaphors with that of novel metaphors, we should use the most conventional metaphoric proverbs in experiments. Earlier I introduced a scale that delineates linguistic meanings and inferences in terms of their degree of conventionality and I said that the meanings of dead metaphors are even more conventional than those of metaphoric proverbs. In the experiments in this thesis we look at the communicability of language-specific metaphoric proverbs, but would it not be better to use dead metaphors as experimental material?

Dead metaphors are unsuitable for testing the competing predictions by CMT, on the one hand, and RT and GS, on the other, concerning the nature of idiomatic meanings of metaphoric expressions. CMT proposes that metaphor comprehension relies on cognitive mechanisms that are general and not specifically linguistic. According to CMT, these mechanisms infer a complex concept that corresponds to the linguistic meaning of a metaphoric expression by way of conceptual metaphors that operate on embodied primary concepts as their semantic constituents. CMT therefore contends that metaphor comprehension is primarily a matter of conception which, according to CMT, supersedes linguistic non-conceptual conventions that might enter the inferential process at some point. RT and GS, contrary to CMT, argue that inferencing

metaphoric meanings is inherently influenced by linguistic conventions that are not conceptual in nature because they are conventions that cannot be said to be part of the conceptual content associated with a metaphor's source and target concepts. RT and GS therefore argue that metaphor comprehension is inherently a combination of linguistic conventions and conception where convention influences *ad hoc* concept derivation, emergent properties and expectations of contextual relevance which in turn affect the way that pragmatic enrichment proceeds and information accumulates pragmatically. On this view these non-conceptual linguistic conventions supersede conception, which is the direct opposite of CMT's prediction.

If we wish to test CMT's and RT and GS' competing predictions experimentally, then we need to study metaphors where both idiomatic and non-idiomatic meanings are available during comprehension. An idiom is a linguistic utterance where the association between its form and meaning is arbitrary because the meaning cannot be inferred compositionally from its constituents. In the case of metaphoric expressions with idiomatic meanings, the constituents are the source and target concepts. If, as CMT claims, construal of the complex concept that corresponds to the idiomatic meaning is primarily a matter of conception operating on the constituent source and target concepts, then the non-idiomatic meanings that speakers infer who are unfamiliar with a proverb's idiomatic meaning should be approximately conceptually similar to its true idiomatic meaning. If, however, RT and GS are correct, then we should be able to find cases where the non-idiomatic meanings that speakers infer who are unfamiliar with a proverb's idiomatic meaning differ from its true idiomatic meaning. This difference, if it exists, should correspond to the contribution of non-conceptual linguistic conventions.

A *dead metaphor* is an idiomatic expression which is processed, we presume, without reference to metaphoric conception. For instance, the idiom *the foot of the hill* may not cause hearers to first think of a FOOT and then metaphorically infer what part of a hill is like a person's foot, they may rely on a conventionalised idiomatic meaning instead. It is important to emphasise that CMT denies the existence of dead metaphors. Instead, CMT argues that metaphoric conception is nevertheless employed, only hearers have become unaware of this metaphoric conception because of the conventionalisation that results from routine use. In the case of the expression *the foot of the hill*, CMT argues that the noun *foot* activates the primary embodied concept FOOT and *hill* activates the embodied concept HILL and metaphor circuitry (metaphoric conception) then matches that part of the hill to FOOT that corresponds to the FOOT's relation to a person's body through analogy. In CMT terms, The metaphoric (idiomatic) meaning of a dead metaphor has become a complex concept which automatically and routinely activates the necessary conceptual metaphors and because conception is automated, there are no conceptually plausible alternative non-idiomatic meanings to be inferred.

The same line of reasoning holds for RT and GS: speakers who are unfamiliar with the conventional meaning of a dead metaphor do not know which conceptual content to retrieve because this content is not *necessarily* made apparent and manifest in the linguistic form of the dead metaphor. *The foot of the hill* might be considered a case where even interlocutors unfamiliar with the conventional meaning can infer it by making use of metaphoric conception. So even when native speakers who are familiar with the conventional meaning of dead metaphors do not make use of metaphoric conception in retrieving the conventional meaning, this does not prevent non-native speakers from doing so. Dead metaphors might be unsuitable to test if there will be differences between the non-idiomatic meanings unfamiliar speakers infer and the idiomatic meanings native speakers infer because the chances for such non-idiomatic alternatives

might be low. If we used dead metaphors as experimental condition, we might risk testing whether speakers are familiar with their conventional meanings, not whether native and non-native speakers make use of similar conception when they infer similar meanings.

3.4 Cross-linguistic metaphor intelligibility

3.4.1 Cross-linguistic approximate conceptual similarity and cross-linguistic metaphor intelligibility

In Chapter 2, Section (2.7), we faced the following controversial and unresolved issue: No two interlocutors may have exactly the same concept in mind for the same expression in natural language yet inferential communication is possible and successful when using that expression. Here I want to reopen the case to what extent two concepts in the minds of two speakers can be approximately similar, how to extend this to a cross-linguistic conceptual comparison, and how to apply this to study cross-linguistic metaphor intelligibility.

If no two speakers can have exactly the same concept for the same natural language expression in their minds, then we might still be able to gauge whether what they have in mind in this case is similar or dissimilar (*gradient similarity*). We might say that two speakers have approximately a similar concept for a natural language expression in mind if what each of them has in mind is more similar than dissimilar and we can say that the approximate similarity is ‘similar enough’ when the similarity between speakers’ two concepts is greater than between either of them and all other concepts in their minds. In other words, two concepts are ‘similar enough’ (approximately conceptually similar) if the similarity between them is greater than between either of them and any and all other concepts.

We can then think of the public lexicon of a language (in RT terms) as an average of the mental lexicons of its speakers. If we nevertheless also account for the fact that individual speakers of this language will have mental lexicons that deviate from the average public lexicon, then we might say that the public lexicons of two languages are significantly different if their sets of averaged concepts, averaged across concepts’ inter-speaker approximate similarity, are more different than the amount by which the mental lexicons of its speakers deviate from the public lexicons within the two languages. We might therefore think of the cross-linguistic similarity of the conceptual systems of two languages as a matter of degree that can be expressed as the likelihood of the mental lexicons of speakers of the two languages to overlap and, although they may never be identical and/or overlap perfectly, we may nevertheless be able to gauge when they are more similar for two given languages than they are for two other languages.

We can then use this notion of approximate similarity of the conceptual systems of two languages to assume that if the conceptual systems of two given languages are more similar than dissimilar and more similar than between any other two languages, then cross-linguistic metaphor intelligibility should be more likely between these two than between any other two languages. If we are looking at a particular set of metaphors, we may gauge cross-linguistic approximate conceptual similarity on a per-metaphor basis rather than by looking at the languages’ two public lexicons as a whole, and we may do the same for the source and target concepts in these metaphors. This is the approach I take in Chapter 4, Section (4.4.4), to establish that (1) the conceptual systems of English and German may be more similar because of language contact than between other languages (Chapter 4, Section 4.4.4), (2) the idiomatic

meanings of proverbs that occur in both English and German are cross-linguistically shared (Chapter 6, Section 6.3), and (3) the source and target concepts of the language-specific and cross-linguistically shared proverbs and the novel metaphors are very similar, both in terms of their linguistic form (Chapter 5) and their semantics (lexical meanings of source and target concepts, compositional meanings, conceptual basis of metaphoric imagery of implied analogy, expectations of contextual relevance; Chapter 6).

3.4.2 Cross-linguistic correlation and non-correlation between the linguistic form and idiomatic meanings of proverbs

Researchers in cognitive science (Lakoff & Johnson, 1980; Lakoff, 1983, 1987; Coulson, 2006) have collected substantial evidence of cross-linguistic systematicity in linguistic metaphors, which they take as evidence for (a) the metaphoric structure of the conceptual systems of the respective languages and (b) the mental representation of the respective linguistic metaphors. They claim that the structure of the conceptual system is inherently metaphoric-analogical and that therefore conceptual interconnections are also metaphor-like. For example, for the complex concept of THOUGHT in English, we might infer the following conceptual metaphors to exist based on the linguistic evidence:

- (6a) THINKING IS MOVING
- (6b) IDEAS ARE LOCATIONS
- (6c) COMMUNICATION IS LEADING
- (6d) UNDERSTANDING IS FOLLOWING

- (7a) THINKING IS PERCEIVING
- (7b) IDEAS ARE THINGS SEEN
- (7c) COMMUNICATION IS SHOWING
- (7d) UNDERSTANDING IS SEEING OR PERCEIVING

- (8a) THINKING IS OBJECT MANIPULATION
- (8b) IDEAS ARE OBJECTS
- (8c) COMMUNICATION IS SENDING
- (8d) UNDERSTANDING IS GRASPING THESE OBJECTS

- (9a) THINKING IS EATING
- (9b) IDEAS ARE FOOD
- (9c) COMMUNICATION IS FEEDING
- (9d) UNDERSTANDING IS DIGESTING

CMT takes the evidence of systematic patterns in linguistic metaphors to be indicative of the kinds of metaphoric conceptual interconnections that exist in the respective languages and these interconnections are therefore called *conceptual metaphors*. On this view, conceptual metaphors

Table 3.1: Cross-linguistic differences in metaphoric proverbs as evidence for differences in conceptual metaphors.

	Similar meaning	Different meaning
Similar metaphor forms	<i>True Positives</i> <i>Rome wasn't built in a day</i>	<i>False Positives</i> <i>to count on the fingers of one's hand</i>
Different metaphor forms	<i>False Negatives</i> <i>frying pan/fire Regen/Traufe</i>	<i>True Negatives</i> <i>A home is a castle Dreams are froth</i>

are *preferred, salient* meanings. They may be universal to us as a species or specific to a particular language, language family, or culture. However, many language-specific linguistic metaphors have conventional meanings called *idiomatic meanings* associated with them. On CMT's view, metaphoric idiomatic meanings are then language-specific conceptual metaphors. Importantly, this characterises idiomatic meanings as language-specific ways of conceptualising perceptual experience and introspective mental states. Contrary to CMT, RT and GS claim that idiomatic meanings are not primarily motivated by conception but by the cooperative need (cf. Grice's Cooperative Principle) to facilitate communication. On CMT's view, metaphoric idiomatic meanings constitute conceptual mental representations. On RT and GS' view, metaphoric idiomatic meanings require knowledge of linguistic conventions that are distinctly *non-conceptual* because they are not primarily motivated by conceptual plausibility, world knowledge, or conceptual knowledge. With respect to metaphoric idiomatic meanings, the difference between the opposing claims made by CMT, on the one hand, and RT and GS, on the other, thus is how much influence metaphoric idiomatic meanings have on the ways that speakers conceptualise the world around them: CMT takes the influence to be rather great and largely unconscious; RT and GS take the influence to be indirectly mitigated by non-conceptual linguistic conventions. We can rephrase this with respect to the evidence of linguistic metaphors: CMT takes the systematicity in linguistic metaphors to be rather direct evidence of the cognitive processes involved and of the mental structures that these processes create; RT and GS take the insight the linguistic evidence gives into cognition to be rather indirect and mitigated by linguistic conventions.

However, we need to be wary of the following cross-linguistic correlations and non-correlations between the linguistic form of proverbs and their metaphoric idiomatic meanings (see Table 3.1). If we want to maximise cross-linguistic approximate conceptual similarity for the idiomatic meanings of metaphoric proverbs, we need to be aware that while the idiomatic meaning might be cross-linguistically shared, the linguistic form might differ. In all of the experiments of this thesis, I require that for all cross-linguistically shared metaphoric proverbs (proverbs that occur both in English and German) they share the same conventional linguistic meaning and the same linguistic form in both English and German, and mental content and representation of the proverbs' source and target concepts are shared between the two languages as well. For all language-specific metaphoric proverbs in the experiments, I require that the source and target concepts are cross-linguistically approximately conceptually similar.

In example (10), we see a metaphoric proverb which has the same wording (linguistic form) in English (10a) and German (10b) and the proverb has the same conventional idiomatic meaning in both languages.

(10a) Rome wasn't built in a day.

(10b) *Rom wurde nicht an einem Tag erbaut.*
 Rome was not on a/one day built.
 'Rome wasn't built in a day.'

Suppose there were non-native speakers of English who did not know whether the metaphor in (10a) had a conventional meaning in English but who, because their native language is German, recognised the wording of the English proverb to resemble the German proverb in (10b). These non-native speakers would thus be justified to assume that the same idiomatic meaning that they know for the German proverb also applies to the English proverb with a similar linguistic form. A cross-linguistic similarity in the linguistic form of a metaphoric proverb thus truthfully and positively correlates with a cross-linguistically similar idiomatic meaning in this case (true positives). Recall that idiomatic meanings act as inferential shortcuts and that speakers should be generally inclined to use inferential shortcuts whenever possible.

However, non-native speakers might also be misled by an apparent cross-linguistic similarity of proverbs' linguistic form. Recall that I defined approximate conceptual similarity as gradient similarity: For two speakers, if, for a particular expression in natural language, the first speaker has concept A in mind and the second speaker has concept B in mind, then A and B are cross-speaker approximately conceptually similar if they are more similar than other concepts that the speakers have (see page 2.7 for a more detailed definition). This should be true because if they are a better match than their conceptual alternatives. When conceptual similarity is maximized, then the communicability of the natural language expression, i.e. speakers' ability to use it to communicate their mental states, should be maximised, too. The example in (11) is precisely a case where there are conceptual alternatives that compete in terms of meaning salience with the correct conventional meanings because they are conceptually similar.

(11a) to be able to count something on the fingers of one's hand

(11b) *an einer Hand abzählen können*
 on a/one hand count-off to can/to be able to
 'to be able to count on one hand' = the true conceptual equivalent of (11a) in German

(11c) *etwas an allen fünf Fingern abzählen können*
 something on all five fingers count-off to can/to be able to
 'to be able to count something on all five fingers (of one's hand)' = false-friend

(11d) to be blindingly obvious / to be clear as day(light) = the true conceptual equivalent of (11c) in English

The English idiom in (11a) and its conceptual German equivalent in (11b) mean that something is a scarce or rare thing, whereas the German idiom in (11c), which resembles (11b) in its linguistic form, means that something is easy to see, comprehend or understand; a distinct difference. Note that the English idiom in (11a) mentions *fingers*. The true cross-linguistic conceptual equivalent of (11a) is (11b). Notice that (11b) does *not* mention *fingers*. The German

idiom in (11c) *does* which does mention *fingers*, conceptually also employs the mental image of counting and therefore is conceptually similar to (11b). However, the idiomatic meaning of (11c), to be blindingly obvious, is not cross-linguistically conceptually equivalent to the idiomatic meaning of the English idiom in (11a), but because (11c) mentions *fingers* same as (11a) does, German non-native speakers of English might falsely assume that the idiomatic meaning of the English expression in (11a) is the idiomatic meaning they know from the German idiom in (11c). The focus of the idiomatic meaning of (11b) is that something is scarce or rare, and, in German, therefore easy to *do*, while the focus of the idiomatic meaning of (11c) is that something is easy to *understand*. So while the idiomatic meaning of (11b) is cross-linguistically more similar to (11a) *conceptually*, the *linguistic form* of (11c) is cross-linguistically more similar to (11a). German non-native speakers of English who are unaware of the correct idiomatic meaning of the English idiom in (11a) might therefore falsely assume that its meaning is that of (11c) and not (11b) because the linguistic form of (11c) is cross-linguistically more similar to (11a). It would thus be wrong of them to assume a cross-linguistic correlation between linguistic form and metaphoric idiomatic meaning. False positive correlations, such as the one in (11c), are also called *false friends*, for that reason. (11c) is the German false friend of the English idiom in (11a).

We might also find cases such as the example in (12), where two languages have proverbs with cross-linguistically approximately conceptually similar idiomatic meanings, but where the linguistic form of the proverbs differ cross-linguistically. The form of the English proverb in (12a) uses the mental imagery of a frying pan and fire to convey the metaphoric meaning that things are going from bad to worse. The German proverb in (12b) expresses the same metaphoric meaning, but instead of the image of a frying pan and fire uses the mental image of rain and heavy rain. Following the rationale employed for the other examples we could naïvely assume that non-native speakers might be misled by the cross-linguistic difference in the form of the proverbs, but I think we would be underestimating speakers' abilities here. Surely speakers of both English and German could make use of their world knowledge to infer that going from a hot frying pan into the fire is not an improvement of one's situation in the same way that getting soaked in a downpour is no better than getting soaked in the rain. So surely speakers of English and German should be able to infer the metaphoric meaning of the proverb from the other language based on their linguistic, conceptual, and world knowledge of the concepts FRYING PAN, FIRE, RAIN, and DOWNPOUR/HEAVY RAIN.

(12a) Out of the frying pan and into the fire.

(12b) *Vom Regen in die Traufe.*
 From the rain (and) into the downpour.
 'Out of the rain and into the downpour.'

But notice that this is possible only because (a) the metaphoric meaning is conceptually plausible based on this knowledge and (b) there are no alternative idioms as in the example in (11). If we are interested in whether non-native speakers are able to infer linguistic meanings for language-specific metaphoric metaphors that are similar to the idiomatic meanings native speakers infer, then we should avoid cases such as (12) because non-native speakers would start the inferential process from source and target concepts that might be different from those of native speakers. They might therefore be misled by the different form and might therefore infer different metaphoric meanings. Therefore, cases such as (12) are a problematic choice

for cross-linguistically shared metaphors. They do not make for a good control condition and should be avoided. Also avoided should be cases such as in example (11), because while cases such as (12) *might* be misleading, cases such as (11) are *highly likely* to be misleading.

And we might find cases such as examples (13) and (14), where there are no metaphoric proverbs in the other language that are cross-linguistically approximately similar, neither in terms of their linguistic form nor their meaning.

(13) A home is a castle.

(14) *Träume sind Schäume.*
Dreams are froth/foam.
'Dreams are froth.'

Language-specific metaphoric proverbs (true negatives) are the test condition in the experiments of this thesis because neither is the linguistic form cross-linguistically approximately similar to alternatives, as in the case of false friends/false positives, nor is the idiomatic meaning cross-linguistically shared but occurs only in one of the two languages but not the other. In preparing the experimental material in Chapter 5, I will require that those metaphoric proverbs categorised as 'language-specific' have neither a cross-linguistically shared linguistic form nor a cross-linguistically shared idiomatic meaning. Cross-linguistically shared metaphoric proverbs (true positives), such as in (10), are one of two control conditions in the experiments (the other control condition being novel metaphors with linguistic meanings that are not yet conventionalised) because monolingual native speakers of both English and German should be familiar with their idiomatic meanings and both the form and idiomatic meanings are cross-linguistically shared.

3.5 Summary

According to RT, interlocutors draw those inferences during inferential comprehension that are cognitively most-relevant to understanding the meaning which the speaker is most likely to imply in a given discourse, as well as understanding the speaker's communicative intent in doing so. Assumptions about which information is cognitively relevant to interlocutors is constantly updated as the discourse progresses. Linguistic meaning is thus highly dependent on the context in which it is uttered, the information gathered from the preceding discourse, interlocutors' conceptual and mental lexicons, and the speaker's communicative intent that the interlocutors infer. RT rejects the Gricean view that comprehending metaphors requires dedicated inferential machinery (Grice, 1975, 1978) and instead claims that there is one set of inferential mechanisms that form the machinery to comprehend any and all linguistic utterances. On this view, inferring a metaphoric meaning differs from other linguistic speaker meanings in the kinds of inferences that have to be drawn in order to satisfy the implied analogy. Pilkington (2000) claims that the inferences drawn in metaphor comprehension are more likely to be a large set of weak implicatures rather than few strong inferences (e.g. entailments or strong implicatures) and these weak implicatures accumulate to produce the rich mental imagery associated with the metaphoric analogy in question.

According to GS, the inferential process consist of two modules: (1) a linguistic module that infers salient meanings based on linguistic knowledge and (2) a pragmatic module that enriches

the meaning by taking non-linguistic knowledge such as conceptual, experiential, perceptual, contextual, and world knowledge into consideration. The more familiar, common, conventional, and prototypical a meaning is, the more likely it is to come to mind more easily and readily, and the more salient and likely to be inferred the meaning is. Inferential strength, on this view, is then a function of *meaning salience*. According to GS, the linguistic meaning that is ultimately inferred for an utterance is one that is a compromise between the outputs of the two modules. The two modules are thought to run in parallel so that processing may align at any point during comprehension. This alignment is necessary in order to optimally arrive at a compromise. On this view, metaphor comprehension requires no dedicated cognitive mechanisms beyond the machinery of the two modules and, similar to RT, that an inferred meaning happens to be metaphoric is then a result of the inferences drawn and not the mechanisms involved. Since the second module allows context to be considered, the linguistic meaning that is inferred by the two modules working in unison is *speaker meaning*. GS' view of metaphoric meanings is thus compatible with that of RT.

CMT claims complex inferences, such as linguistic meanings, are the result of conceptual mappings from primary embodied concepts to ever more complex higher-order concepts. Which mappings are established and activated to yield complex inferences, according to CMT, depends on sensory-perceptual stimuli (context, in RT and GS terms) and the mental content and structure of concepts based on past experiences. Although CMT thinks of linguistic meanings as the result of interconnecting primary concepts to ever more complex higher-order concepts, this view is compatible with RT and GS' view of *speaker meaning*. The complex inferences that complex higher-order concepts yield for a particular context, in CMT terms, are the complex inferences that correspond to context-dependent speaker meanings, in RT and GS terms. However, different from RT and GS, CMT takes metaphoric-analogical reasoning to be a structural feature of the mental architecture of the conceptual system. Comprehending metaphors, on this view, thus comes natural to the human mind. However, this position also commits CMT to claim that metaphoric idiomatic meanings are part of speakers' conceptual knowledge and are ways of conceptualising. On this view, conventionalisation of metaphoric meanings to idiomatic meanings is the process of strengthening and solidifying language-specific conceptual metaphors. CMT, RT and GS thus disagree what it is that is conventionalised in metaphoric idiomatic meanings. For CMT, conceptual metaphors are conventionalised. For RT and GS, the conventions are non-conceptual constraints on linguistic expression of mental imagery.

CMT assumes that if inferential comprehension is primarily a conceptual process, then when speakers have similar concepts pertaining to metaphors' language-specific idiomatic meanings (*A*), they should also comprehend these metaphors similarly (*B*). CMT raises the stakes of what it means to assume $A \rightarrow B$ because for CMT we are not just making an assumption about the cognitive mechanisms of metaphor comprehension, we are making an assumption about analytical reasoning in general. So while RT, GS, and CMT are interested in understanding the *cognitive* mechanisms underlying inferential metaphor comprehension, CMT risks more by claiming that the mechanisms are not dedicated to language but general mechanisms of analogical reasoning. Ockham's razor would suggest that if $A \rightarrow B$, CMT offers a simpler, more general, more appealing theory of metaphor comprehension than RT and GS. RT, GS, and CMT agree that conventional idiomatic metaphoric meanings act as inferential shortcuts: instead of having to consider many potential interpretations of the implied analogy and decid-

ing which one is the most plausible in the context of a given metaphor, speakers can directly default to the idiomatic meaning. According to CMT, idiomatic metaphoric meanings are sets of language-specific conceptual metaphors. According to RT and GS, they are language-specific conventions of conceptualising that are not motivated by or may even go against pure conceptual plausibility, but are instead motivated by the need to facilitate effective communication. So for CMT, idiomatic metaphoric meanings are conceptual; for RT and GS, they are non-conceptual constraints on conception. We can therefore rephrase the question of whether CMT's assumption is true by asking whether conceptual or non-conceptual aspects of idiomatic metaphoric meanings are more important in how speakers infer these meanings. The validity of CMT's assumption that $A \rightarrow B$ depends on the condition that conceptual cross-linguistic similarity is maximised for the metaphors and speakers in question. The rephrased question depends on the same condition, so testing the validity of the rephrased question is also a test of CMT's assumption that metaphor comprehension is primarily conceptual.

At the end of this chapter we saw that we need to be wary of ambiguous non-correlations between metaphors' linguistic form and idiomatic meanings and avoid them when selecting metaphors with language-specific idiomatic meanings for experimentation. In the next chapter, Chapter 4, I argue that the closeness of language contact makes English and German an ideal choice to test CMT, RT and GS' competing assumptions because cross-linguistic conceptual similarity should be *more likely* between English and German than other languages. Chapter 4 also discusses how we may operationalise the theoretical notions introduced in this and the previous chapter for experimentation: conceptual cross-linguistic similarity, meaning salience of the concepts necessary to infer the idiomatic meaning, cross-linguistically similar expectations of contextual relevance, conventionality of a metaphor's linguistic form, meaning, and form-meaning association.

Chapter 4

Methodology: cross-linguistic metaphor intelligibility

4.1 Introduction

In this chapter I lay out the central research questions, the fundamental definitions pertaining to the experiments of this research and their operationalisation for experimental measurement. In Section (4.2), I introduce the central questions of this research. Section (4.3) describes how metaphoric proverbs can be selected for experiments such that they are conceptually similar between English and German. In Section (4.4), I describe how I think the theoretical notions pertaining to the research questions should be operationalised in such a way that they can be measured experimentally, particularly two ways of operationalising cross-linguistic intelligibility (weak and strong intelligibility). This section also discusses the three ways (measurands) in which cross-linguistic metaphor intelligibility is measured (4.4.1): (1) reading/response times, (2) plausibility judgements, and (3) a context creation task. In Section (4.4.2), I define the criteria that speakers need to meet in order to be categorised as monolinguals and bilinguals. Section (4.4.4) describes how we can quantify the closeness of linguistic contact between English and German, and in the last section, Section (4.4.5), I define the notion of language-specificity of metaphors employed in the experiments.

4.2 Research questions

Researchers in cognitive science (Lakoff & Johnson, 1980; Lakoff, 1983, 1987; Coulson, 2006) have collected substantial evidence of cross-linguistic systematicity in linguistic metaphors, which they take as evidence for (a) the metaphoric structure of the conceptual systems of the respective languages and (b) the mental representation of the respective linguistic metaphors. They claim that the structure of the conceptual system is inherently metaphoric-analogical and that therefore conceptual interconnections are also metaphor-like. The evidence of systematic patterns in linguistic metaphors, on this view, is taken to be indicative of the kinds of metaphoric conceptual interconnections that exist in the respective languages and these interconnections are therefore called *conceptual metaphors*. On this view, conceptual metaphors are preferred, salient meanings. They may be universal to us as a species or specific to a particular

language, language family, or culture. However, many language-specific linguistic metaphors have conventional meanings called *idiomatic meanings* associated with them. On the view held by Conceptual Metaphor Theory (CMT) (Lakoff & Johnson, 1980; Lakoff, 1983, 1987; Lakoff & Johnson, 1999, 2003; Lakoff, 2008, 2009), metaphoric idiomatic meanings are then language-specific conceptual metaphors. Importantly, this characterises idiomatic meanings as language-specific ways of conceptualising perceptual experience and introspective mental states. Contrary to CMT, Relevance Theory (RT) (Sperber & Wilson, 1986; Wilson & Sperber, 1993; Sperber & Wilson, 1995; Wilson & Sperber, 2002, 2004) and Graded Salience (GS) (Giora, 1997, 1998, 1999, 2003; Peleg et al., 2008; Peleg & Giora, 2011) claims that idiomatic meanings are not primarily motivated by conception but by the cooperative need (cf. Grice's Cooperative Principle) to facilitate communication. On CMT's view, metaphoric idiomatic meanings constitute conceptual mental representations. On RT and GS's view, metaphoric idiomatic meanings require knowledge of linguistic conventions that are distinctly *non-conceptual* because they are not primarily motivated by conceptual plausibility, world knowledge, or conceptual knowledge. With respect to metaphoric idiomatic meanings, the difference between the opposing claims made by CMT, on the one hand, and RT and GS, on the other, thus is how much influence metaphoric idiomatic meanings have on the ways that speakers conceptualise the world around them: CMT takes the influence to be rather great and largely unconscious; RT and GS take the influence to be indirectly mitigated by non-conceptual linguistic conventions. We can rephrase this with respect to the evidence of linguistic metaphors: CMT takes the systematicity in linguistic metaphors to be rather direct evidence of the cognitive processes involved and of the mental structures that these processes create; RT and GS take the insight the linguistic evidence gives into cognition to be rather indirect and mitigated by linguistic conventions.

The less contact there is between two languages and cultures, the more likely it should be to find conceptual differences if there are any. In the previous chapter, I discussed CMT's claims that the underlying mechanisms of thought are culturally-transmitted conceptual mappings called *conceptual metaphors* that can be combined to facilitate the inferential comprehension of more complex concepts. CMT claims that these conceptual mappings are similar in nature to metaphoric analogies. Since CMT claims these conceptual mappings are culturally-transmitted, cultures with intense linguistic contact would exchange concepts along with these conceptual mappings. If indeed these conceptual mappings form the basis for comprehending more complex ideas, especially those that require one to draw metaphoric analogies during comprehension, then two cultures in close linguistic contact should be unlikely candidates for finding conceptual differences. The empirical literature focusses on cultures with little linguistic contact, but I think we also need to consider languages in close contact: English and German are two languages that arguably have one of the closest language contacts among the languages of the world. This thesis thus asks:

- (1) Are metaphors from one language intelligible to speakers of another language if those two languages are in close linguistic contact and thus can be assumed to share similar or the same concepts?

If CMT is correct in its predictions and comprehension of linguistic metaphors requires metaphor-like mappings within the conceptual system of the mind, then the conceptual metaphors necessary for comprehending specific linguistic metaphors should be available to speakers of both languages where the two languages are in close linguistic contact and speakers grow up with

exposure to the same basic conceptual metaphors. The linguistic theories, on the other hand, would predict contrary to CMT that there might be linguistic metaphors in the two languages that have language-specific idiomatic meanings that are not inherently cross-linguistically intelligible to speakers of the other language. And for linguistic metaphors with idiomatic meanings this should be the case regardless of whether speakers of both languages conceptualise the world in the same way because the idiomatic meanings are tied into speakers' language-specific grammar through *sense relations* that are not part of or subject to the conceptual system. Sense relations are connected to concepts, i.e. they map onto the conceptual system just as the conceptual system maps onto them, but, importantly, not necessarily in a one-to-one fashion in each particular case (e.g. English *pig* for the animal and *pork* for its meat, French *porc* for both the animal and its meat). Conversely, it should not matter for the predictions of CMT whether specific linguistic metaphors have idiomatic meanings or not because in all cases their concepts and conceptual mappings should be rooted in and built from basic conceptual mappings and to say that a linguistic metaphor has an *idiomatic* meaning, in CMT, should mean merely that the neural connections that embody the composition of the concepts and conceptual mappings necessary for that metaphor from basic conceptual mappings have been *recruited* so many times through cultural exposure that they become permanent (*neural binding*); in other words, the idiomatic meaning is conventionalised, culturally transmitted, and learned by each generation of speakers.

It is well-known, especially through the contributions of RT, that *speaker meaning* largely depends on the context in which an utterance is made. CMT and linguistic theories of inferential comprehension agree that contextual information is relevant if it yields positive cognitive effects; in other words, if it is informative. It is informative if it contributes something new to the discourse, aids comprehension, or is an explanation of conceptual metaphoric mappings, i.e. if it is a substitute for the conceptual metaphors that speakers might lack. Context should make it easier to form *ad hoc* concepts. In other words, context is relevant if it is helpful during comprehension. If contextual information is relevant for comprehending a metaphor, then speakers should make use of it whenever possible. If contextual information is irrelevant for comprehending a metaphor or conceptually conflicts with what speakers already inferred, speakers should ignore this new contextual information. If speakers come from different language communities and cultures, linguistic metaphors such as metaphoric proverbs that have language-specific meanings might be cross-linguistically unintelligible to them without sufficient context. With relevant contextual information, however, cross-linguistically unintelligible metaphors should become intelligible. This research therefore also asks a follow-up question to the first question:

- (2) If context improves the intelligibility of metaphors across languages, do bilinguals make use of context in the same way that monolinguals do?

In this second question, we look at potential differences between monolinguals and bilinguals because bilinguals can be understood to be members of both of the language communities and cultures that we are comparing in the first question, while monolinguals are only members of one language and culture but not the other.

In this thesis I argue that idiomatic metaphoric proverbs can be identified for English and German that do not occur in the other language and that are cross-linguistically less intelligible to speakers of the other language than metaphoric proverbs from their native language. In

my investigation I do not *assume* that English and German do share similar or the same concepts but *show* that they do for the metaphoric proverbs in question by critically assessing the proverbs' cross-linguistic similarity in terms of their semantics and linguistic form. I also look at the familiarity of the linguistic form as a measure of conventionality. I find that the form, familiarity, and semantics of the proverbs is cross-linguistically conceptually similar and I argue that the reason why the English and German metaphoric proverbs are less intelligible to native speakers of the other language than metaphors from their native language is *not* because of differences in the kinds of conceptual mappings available to speakers but because of differences in linguistic knowledge, i.e. *sense relations* and *expectations of contextual relevance*. If this interpretation is true then this thesis provides empirical counterevidence to CMT found in the unlikeliest of places.

4.3 Toward a methodology to maximise the chance of cross-linguistic metaphor intelligibility

Recall that in testing the predictions made by CMT, on the one hand, and RT and GS, on the other, we are trying to maximise cross-linguistic approximate conceptual similarity on a per-metaphor basis in order to maximise the likelihood of cross-linguistic metaphor intelligibility of language-specific proverbs and we measure this metaphor intelligibility by gauging the difference in mental content between the non-idiomatic meanings that speakers who are unfamiliar with proverbs' idiomatic meanings might draw and the true idiomatic meanings. If CMT's predictions are correct, this difference should be small or non-existent (in the sense of approximate conceptual similarity). If RT and GS are correct, this difference should be significantly large (in the sense of approximate conceptual similarity) and larger than what we would expect if CMT was correct. The research question is: When non-native speakers see language-specific metaphoric proverbs from the other language (translated into their native language), will they infer the same meaning that native speakers have through linguistic convention (the idiomatic meaning) if the two languages are in close linguistic contact?

It makes sense to compare non-native speakers' performance on metaphoric proverbs with language-specific idiomatic meanings to those with cross-linguistically shared ones because we already know that native and non-native speakers will know the idiomatic meanings of cross-linguistically shared proverbs. We should therefore choose cross-linguistically shared metaphoric proverbs as a control condition, i.e. proverbs that share the same idiomatic meaning for two languages whose public lexicons are cross-linguistically conceptually more similar than either of them and any other language. However, because we are measuring cross-linguistic metaphor intelligibility by gauging the difference in inferential content in non-idiomatic and idiomatic meanings, we need to choose only those metaphoric proverbs with cross-linguistically shared idiomatic meanings that also share the same linguistic form because if they did not, then I find it questionable whether we can say that the mental content of the "shared" idiomatic meaning is truly the same if the two languages employ different metaphoric mental imagery to evoke it.

In the example of *out of the frying pan and into the fire*/*out of the rain and into the downpour* from the previous chapter, the English and German proverb express the same sentiment and this sentiment can be inferred from either one of the proverbs: the frying pan is hot; if there is enough heat, then being in the frying pan is potentially dangerous; in order to evade this

danger, you would want to escape from the frying pan. But by doing so we would land in the fire, which is as hot as the frying pan if not more so, which is not an improvement over the predicament we are in. Although our intentions were to avoid danger, the situation has therefore gone from bad to worse. Getting soaked in the rain is not pleasant and if you were exposed to wet and cold weather for too long it might become hazardous to your health; we therefore would want to avoid the rain. But if in trying to avoid the rain, we get into even heavier rain, we will be even wetter and the danger to our health increases even more. Therefore the situation we are in has gone from bad to worse. So while the ultimate conclusions we draw when inferring the meaning of the two proverbs are the same (that in an attempt to improve one's lot, the situation has actually gone from bad to worse) and this meaning is cross-linguistically shared, the metaphoric imagery that implies this sentiment differs between English and German. This difference in mental imagery, I think, would have to be reflected in the mental content of the proverb's idiomatic meaning for English and German speakers. I am therefore not convinced that we can say that for metaphoric proverbs such as these, where there is a non-correlation between their linguistic form and idiomatic meaning, we can say the idiomatic meanings are cross-linguistically shared (not even in the sense of cross-linguistic approximate conceptual similarity). In the experiments of this thesis I therefore avoid proverbs with a cross-linguistic non-correlation between form and meaning and only use proverbs with the same cross-linguistically shared idiomatic meaning and the same linguistic form as a control condition.

Examples of non-correlation are not in itself evidence against CMT. Recall, CMT takes complex concepts, especially abstract ones, to be the result of interconnecting many primary embodied concepts. The notion of a situation or set of personal circumstances going from bad to worse is an abstract concept. When asked we could give many examples that could represent this idea metaphorically. The cross-linguistic non-correlation between English and German for this specific proverb simply shows that English speakers have a different set of embodied concepts that, so CMT could argue, interconnect to the metaphor concept and German speakers have another set of embodied concepts. The English set of embodied concepts then gives rise to the mental imagery and this imagery differs from the imagery that the German set of embodied concepts give rise to. There is nothing about two language-specific sets of mental imagery leading to (cross-linguistically) the same complex metaphor concept. On the contrary, given CMT's claim about the nature of abstract concepts, we should expect such examples of false positives in the metaphor domain. However, we should avoid false positives when we pick metaphoric proverbs with language-specific idiomatic meanings. In the experiments, metaphors' language specificity is the specificity of their idiomatic meanings. So, strictly speaking, *out of the frying pan and into the fire*/*out of the rain and into the downpour* would have to be classified as cross-linguistically shared metaphors, but because their linguistic form differs cross-linguistically, English and German participants might think they were language-specific. However, what we want are metaphors whose idiomatic meanings are *unambiguously* language-specific or cross-linguistically shared. Ambiguity in the association between linguistic form and meaning thus needs to be avoided.

L2 metaphors are true negatives, not false negatives because there is no proverb in the L1 with the same metaphoric meaning as the idiomatic meaning in the L2. For false negatives, there is a proverb in the L1 *with* a conventionally established idiomatic meaning. For the L2 metaphors selected for the experiments, there is not. This is why only the L2 metaphors

of the experiments allow us to see whether speakers' L1 conceptual knowledge pertaining to metaphors' source and target concepts and their compositional metaphoric meaning are enough to facilitate inferring meanings such as are conventionalised as idiomatic ones. Only this way can we test how much of idiomatic metaphoric meanings is conceptual in speakers' minds and how much is non-conceptual. If we used false negatives in the experiments instead of true negatives, we would be testing speakers' recall of idiomatic metaphors that are most similar to the L2 metaphors semantically. We would not know any more than before about how much conceptual and non-conceptual knowledge factors into idiomatic meanings that have such a strong influence on native speakers' intuitions of meaning salience. For this reason, the question could not be settled theoretically without experimentation. Also, to generally favour theory over experiment leads to bad scientific practice that is really bad at spotting its own mistaken assumptions. Just because something makes sense in theory, no matter how rigorously proved, does not mean that nature, the mind, its psychology and physiology, work like that.

I think we can maximise the chance of cross-linguistic metaphor communicability and intelligibility by maximising the chance of cross-linguistic approximate conceptual similarity on a per-metaphor, per-source, and per-target concept basis. If cross-linguistic approximate conceptual similarity is maximised for a particular set of metaphoric proverbs with language-specific idiomatic meanings, then we should expect that speakers of two languages in close linguistic contact should find these language-specific proverbs as intelligible as proverbs with the same idiomatic meanings in both languages. This is the assumption made by CMT. This assumption, however, presupposes that cross-linguistic metaphor intelligibility is primarily a matter of conception and not a matter of having knowledge of language-specific non-conceptual linguistic conventions. CMT takes the evidence of metaphor systematicity within and across languages to be indicative of the conceptual systems of the respective languages, i.e. it takes metaphor systematicity to be indicative of the way that speakers of these languages conceptualise each other and the world around them. CMT also claims that this mode of conception, which linguistic metaphors are supposedly indicative of, takes precedence over knowledge of linguistic conventions, and it assumes that most of this metaphoric conception (*conceptual metaphors*) is unconscious while knowledge of linguistic conventions is more conscious (Lakoff, 2008).

RT disagrees with CMT's assumption that knowledge of linguistic conventions is not unconscious. Most relevance theorists think of the inferential comprehension process as guided by cognitive constraints—a focus on contextually relevant information that pertains to the current discourse, constraints on the amount of cognitive resources spent during processing and for a given task, and constraints on the depth of inferencing and the preferences given to drawing particular inferences rather than others in a given conversational context—and they think of these constraints as not necessarily conscious or subject to executive control or rational decision making. I agree with RT here and therefore do not find CMT's assumption that metaphoric conception necessarily is more unconscious than linguistic knowledge of conventions convincing and thus reject this assumption. Nevertheless, the question whether metaphoric conception or knowledge of non-conceptual linguistic convention is more important in inferring the idiomatic meaning of metaphoric proverbs still stands, I think, even if one rejects CMT's assumptions about consciousness. The reason why I think this is because even if both metaphoric conception and knowledge of non-conceptual linguistic convention were equally unconscious (or equally conscious, for that matter), one might still play a more important role during metaphor comprehension and I think we cannot decide this issue *a priori* and should therefore try to

test it experimentally. I therefore reject CMT's argument that the level of consciousness is the cause of supposedly making metaphoric conception more primary than knowledge of linguistic convention, but at the same time I recognise that CMT takes metaphoric conception to be more primary during inferencing while RT and GS take knowledge of linguistic conventions to be more important than metaphoric conception. I think we therefore have an opportunity here to test the predictions concerning primacy made by CMT, on the one hand, against RT and GS, on the other.

CMT claims the idiomatic meaning of a particular metaphoric expression, such as a language-specific metaphoric proverb, is a complex concept and inferring a complex concept from its constituent embodied primary concepts (i.e. the source and target concepts of the metaphoric expression) is achieved through conceptual metaphors. Furthermore, CMT claims that establishing particular conceptual metaphors is motivated by exposure to particular sensory stimuli and experiences, in CMT terminology, i.e. context and world knowledge in RT and GS terminology. Given these claims, CMT would predict that if two languages share all the primary concepts and conceptual metaphors (in the sense of cross-linguistic approximate conceptual similarity) necessary to infer the meaning of a particular metaphoric proverb, then non-native speakers should be able to infer the same complex concept (the metaphoric meaning, in the sense of approximate conceptual similarity, "similar enough") led by conceptual plausibility. This prediction amounts to saying that non-native speakers need only consider their conceptual knowledge to draw plausible inferences about metaphoric expressions which would be similar to the inferences that native speakers draw, and that although the metaphoric expressions are unfamiliar and new to the non-native speakers.

If the idiomatic meaning of a metaphoric expression, in RT and GS terms, is a linguistic meaning that is inferred on the basis of the source and target concept and based on the given context or a context that is associated out of considerations of inferential-cognitive relevance, but, critically, this linguistic meaning is not inferable without knowledge of non-conceptual linguistic conventions, conventions that are not motivated by or subject to conceptual plausibility, then speakers who are unaware and unfamiliar with these non-conceptual conventions should be unable to infer meanings for metaphoric expressions that are approximately similar to their idiomatic meanings. In other words, if idiomatic meanings require knowledge of linguistic conventions that have nothing to do with how speakers conceptualise but which only constrain how they communicate these ideas (concepts), then in a cross-linguistic situation where we know that speakers of two language communities conceptualise in similar ways (with respect to the concepts of particular metaphoric expressions) speakers who do not know the relevant L2 language-specific linguistic conventions should infer metaphoric meanings that are distinctly and significantly different from the true idiomatic meanings that speaker familiar with the conventions infer. If idiomatic meanings require knowledge of particular linguistic conventions, then when speakers do not have this knowledge, they should be unable to infer these idiomatic meanings, i.e. they should infer other plausible linguistic meanings.

If we set up a situation where the chances of cross-linguistic intelligibility of metaphoric proverbs are maximised conceptually and if, as CMT proposes, cross-linguistic metaphor comprehension is primarily dependent on cross-linguistic approximate conceptual similarity, then language-specific metaphoric proverbs should be cross-linguistically as intelligible as cross-linguistically shared proverbs. If, however, cross-linguistic metaphor comprehension is not primarily dependent on metaphoric conception but approximate similarity of non-conceptual

linguistic conventions concerning proverbs' idiomatic meanings, as RT and GS propose, then cross-linguistic intelligibility should be lower for language-specific metaphoric proverbs than for cross-linguistically shared ones even when cross-linguistic approximate conceptual similarity is maximised for all metaphors in question.

Because of the notion of *conceptual metaphor* in CMT, we may rephrase the above research question as: are idiomatic meanings of metaphoric proverbs language-specific conceptual metaphors, i.e. cross-domain, cross-modal cognitive associations motivated by conceptual plausibility in the framework of Embodied Cognition, or are idiomatic meanings language-specific non-conceptual linguistic conventions, i.e. conventions of conceptualising that are not motivated by or may even go against pure conceptual plausibility? This rephrased version of the research question assumes that plausibility correlates with and can therefore be used as an indicator of metaphor intelligibility, which in turn is an indicator of metaphor communicability: in order for a metaphor to be communicative it needs to be intelligible. In order to be intelligible a metaphor needs to be plausible. A metaphor that is not plausible, I propose we assume here, cannot be intelligible and therefore cannot be communicative.

In the experiments of this thesis, we will measure metaphors' plausibility, which is effectively a measure of *metaphor aptness*. The most widely accepted definition of metaphor aptness defines it as "the extent to which the statement captures important features of the topic" (Chiappe, Kennedy & Smykowski, 2007: 97). Jones & Estes (2006) argue that metaphor aptness and conventionality are highly correlated. I therefore take it to be of central concern for our investigation to establish an objective measure of metaphor conventionality and familiarity. Subjective measures of conventionality and familiarity ask experimental participants to rate metaphors accordingly (see, for instance, Blasko & Connine, 1993). Objective measures of conventionality and familiarity try to establish speakers' likelihood to be familiar with a given metaphor *independent* of speakers' introspective judgement. Thibodeau & Durgin (2011) advocate corpus frequency as an objective measure of metaphor familiarity. They used the Google online search as a corpus. This thesis uses part-of-speech tagged (BNC, COCA, COHA, COSMAS-II) and n-gram Google corpora.

4.3.1 Splitting metaphor conventionality into familiarity with proverbs' linguistic form and the arbitrariness of the form-meaning association

The primary rationale of the experiments in this thesis is to maximise cross-linguistic approximate conceptual similarity for a set of metaphors so that cross-linguistic metaphor intelligibility should be highly likely and to look for cases of lowered cross-linguistic metaphor intelligibility. In order to test the competing predictions of CMT, on the one hand, and RT and GS, on the other, we want language-specific metaphoric proverbs with highly conventional idiomatic meanings. We therefore need a way to determine the *degree of conventionality* of metaphoric expressions. In order to determine whether the conventionality of proverbs' idiomatic meanings is maximised we need to be able to quantify conventionality. A metaphoric proverb with an idiomatic meaning is a linguistic metaphor where the association between its linguistic form and meaning is arbitrary because the meaning cannot be inferred compositionally from its constituent source and target concepts. The aspects that make a metaphoric proverb with an idiomatic meaning conventional are (1) the conventionality of its particular wording (lin-

guistic form), (2) the salience of the idiomatic meaning, and (3) the conventionality of the association between its linguistic form and idiomatic meaning. I therefore propose that we split the notion of conventionality into two notions, one pertaining to the linguistic form of a metaphoric proverb (the familiarity with a proverb's linguistic form) and the other pertaining to its linguistic meaning (the arbitrariness of the form-meaning association).

4.3.2 Corpus frequency as an index of familiarity with proverbs' form

Speakers' familiarity with a proverb's linguistic form is part of what determines that proverb's conventionality: The more speakers are familiar with a particular linguistic form of a proverb which they associate with a particular idiomatic meaning, the more conventional the form and meaning association of it is. At the same time, the more the individual speaker of a language is exposed to instances of a particular proverb in a recurring form, the more likely that speaker is to acquire that proverb and to strongly associate the most common linguistic form used with its idiomatic meaning with that proverb. The more frequent a metaphoric proverb is in speakers' individual exposure to the language, the more likely they are to familiarise themselves with its form, idiomatic meaning, and conventional arbitrary form-meaning association. The more frequently a proverb occurs in a language overall, the more likely individual speakers are to familiarise themselves with it and the more speakers are likely to become familiar with it. The more speakers familiarise themselves with a particular proverb, the more likely it becomes that they will start using it themselves, which further increases the chance of other speakers who have not yet encountered the proverb to encounter it and to familiarise themselves with it. Frequency of proverb use thus gives us a measure of the chance of speakers of a language community to be familiar with it. At the same time, although we cannot be certain that if we pick a speaker at random from the population of the language community that that speaker will be familiar with a particular proverb, we can nevertheless make general predictions about the chances of the average speaker (of being familiar with the proverb).

Corpora are one source to gauge the usage frequency of a particular metaphoric proverb. Corpora that are larger and that sample from a variety of modes of language production—such as different text genres of written language, transcriptions of spoken language, spontaneous, rehearsed, institutional, or circumstantial speech—are more representative of the actual usage of a proverb than corpora which are smaller in size and that sample only particular modes of language production. The more representative a particular corpus is, the more representative the frequency estimates of a proverb's actual use are and the more reliable the impression of speakers' likelihood to be familiar with it is that we gather from the corpus. Since corpora record linguistic forms and not proverbs' idiomatic meanings, the estimates of usage frequency and familiarity likelihood are indicators of speakers' familiarity with proverbs' form. The more frequently a proverb is expressed in a particular linguistic form, the more likely speakers are to be familiar with this form rather than alternative forms and the more likely it is that the most frequently used form is also the most conventional one.

4.3.3 Conceptual plausibility as an index of the arbitrariness of the form-meaning association

What makes the meaning conventionally associated with a particular proverb idiomatic is that the association between the linguistic form of the proverb and this meaning is arbitrary in

the sense that the meaning is not inferable from its form. The meaning of an utterance is inferable from its form if it is inferable from the compositional semantics of its constituent (lexical) concepts, the world knowledge in the mental lexicons of individual speakers associated with the constituent concepts, and/or the discourse context in which the form is used given that that context contains information relevant to the comprehension of the form in question (relevant in RT terms). A metaphoric expression consists of a source and target concept which, assuming that they are both overtly expressed in the linguistic form (recall, target concepts may be covert), are minimally connected syntactically by a copula form of the verb BE, as in *time is money*, i.e. the source concept MONEY is syntactically attributed to the target concept TIME. This direct attribution that the syntax suggests, however, fails semantically but can be rescued by deriving an ad hoc concept from the original source concept MONEY whose content has been pragmatically enriched such as to facilitate the analogy implied between the source and target concept. The idiomatic meaning is not inferable if and when its content cannot be inferred through indirect semantic attribution via derivation of ad hoc concepts and pragmatic enrichment. Nevertheless, other meanings may be inferable from proverbs' from but those meanings must not have the same content as the idiomatic meaning. The inferences drawn in order to facilitate the implied analogy during ad hoc concept derivation and pragmatic enrichment vary in inferential strength from highly salient (entailments and possibly strong implicatures) to lowly salient (weakest implicatures). The stronger the inferential strength and salience of inferences, the more justification based on context, world and conceptual knowledge there is for drawing these inferences. Meanings that can be inferred from the linguistic form of a metaphoric proverb (i.e. the ones that would be non-idiomatic meanings) are more plausible in the eyes of speakers, the more justification there is for drawing the necessary inferences and drawing particular inferences is more justified if the inferences are more salient. A proverb's idiomatic meaning is highly salient, but because it is an inferential default interpretation given by linguistic convention and not because it was inferable. Thus, the form-meaning association for the idiomatic meaning is between the linguistic form of the proverb as a whole, whereas alternative non-idiomatic meanings, even if they are as salient as the idiomatic meaning, are inferred from the constituent source and target concepts and the implied analogy between them. Hence, the more salient non-idiomatic meanings which satisfy the conditions of the implied analogy (indirect attribution) are, the more conceptually plausible they are, i.e. plausible based on the mental content of the source and target concept and knowledge associated with them. Therefore, the more the content of a proverb's idiomatic meaning differs from the content of inferable salient non-idiomatic alternative meanings, the more arbitrary the form-meaning association between the linguistic form of a proverb and its idiomatic meaning is.

Note that native speakers may consider the association between the form of a proverb and its idiomatic meaning to be conceptually plausible, but thus intuition might be a result of the linguistic convention that enforces the idiomatic meaning. By choosing two languages where the source and target concepts of the proverbs under investigation cross-linguistically have approximately similar content, more similar than for other language pairings, non-native speakers who are unfamiliar with the idiomatic meaning cannot use the inferential shortcut but have to infer a salient (justified) and plausible non-idiomatic meaning for L2 language-specific proverbs. We can thus gauge the difference between the content of the non-idiomatic meanings that non-native speakers draw and compare them to what native speakers consider to be the content of proverbs' idiomatic meanings. If so, then the more arbitrary the association

between the linguistic form and idiomatic meaning of a language-specific proverb is, the more the content of non-native speakers' inferred non-idiomatic meanings should differ from native speakers' idiomatic meanings. If a proverb's idiomatic meaning is thus not inferable from the constituent source and target concept, i.e. judged on the basis of conceptually plausible inferences alone, then the association between its form and idiomatic meaning will be arbitrary. The more arbitrary the association between the linguistic form of language-specific proverbs and their idiomatic meanings is, the more implausible and incomprehensible language-specific proverbs should be to non-native speakers. Hence, I propose that non-native speakers' intuitions of conceptual plausibility are a better gauge of the arbitrariness of the association between the form and idiomatic meaning of language-specific proverbs than the intuitions of native speakers if we independently maximise the cross-linguistic approximate conceptual similarity of the source and target concepts of the metaphoric proverbs. Non-native speakers' intuitions of conceptual plausibility might be a better gauge because native speakers might not be able to dissociate those parts of their inferential process of the idiomatic meaning pertaining to the content of the source and target concept and those parts that result from following the linguistic convention that constrains their interpretation of the linguistic form of the proverb as a whole. Non-native speakers, because they are unaware of and unfamiliar with the linguistic conventions that might influence native speakers, are in a position where they are forced to judge the conceptual plausibility of the pragmatically enriched implied analogy based on their knowledge pertaining to proverbs' source and target concepts alone. Hence, non-native speakers' intuitions of conceptual plausibility are more likely to be based on metaphors' source and target concepts, while native speakers' intuitions of conceptual plausibility are more likely to relate to the linguistic conventions that constrain proverbs' idiomatic meanings. I propose that we operationalise the arbitrariness of the association between proverbs' form and idiomatic meaning by assuming that differences in native and non-native speakers' perceived conceptual plausibility of the metaphoric proverbs should correlate with the arbitrariness of the association between proverbs' forms and idiomatic meanings. The more arbitrary the form-meaning association is, the greater the differences in perceived conceptual plausibility between native and non-native speakers should be.

If, as CMT proposes, idiomatic meanings could be considered conceptual metaphors and conceptual metaphors are conceptually plausible mappings in the mind because they are motivated by our shared distinctly human embodied cognition and common conception correlates with a common culture, then the arbitrariness of the association between metaphoric proverbs' form and idiomatic meanings should be small as judged by non-native speakers for a pair of languages where the source and target concepts are cross-linguistically approximately conceptually similar and where speakers of both languages are likely to have similar conceptual metaphors in general because of a common cultural background. If, however, as RT and GS propose, it is possible that the linguistic conventions constraining idiomatic meanings are not inherently motivated by conceptual plausibility, then we should expect to find metaphoric language-specific proverbs where the association between their linguistic form and idiomatic meanings is arbitrary, as judged by non-native speakers, even for a pair of languages where the source and target concepts of the proverbs are cross-linguistically approximately conceptually similar and the languages have the same cultural background.

Once we have an operationalised measure of the arbitrariness of the association between the linguistic form of language-specific proverbs and their idiomatic meanings, we can then assume

that the arbitrariness of the form-meaning association should correlate with conventionality: the more arbitrary the form-meaning association, the more conventional the idiomatic meaning of the proverb in question is. Note that I focus on the conventionality of the idiomatic meaning here. For the conventionality of the linguistic form of the proverb I proposed to assume a correlation between the conventionality of the form and speakers' familiarity with that form. The measures of familiarity and arbitrariness of the form-meaning association thus complement each other to give us a more complete picture of the notion of conventionality of language-specific metaphoric proverbs in order to be able to gauge and manipulate the magnitude of cross-linguistic conceptual similarity, salience, familiarity, and arbitrariness to test their effects on the cross-linguistic intelligibility of these language-specific proverbs.

4.4 Definitions and operationalisations

4.4.1 Cross-linguistic intelligibility

Metaphors are communicative if they are intelligible. We need to figure out how to operationalise the measurand *cross-linguistic intelligibility*, i.e. we need to find a way to define cross-linguistic intelligibility in such a way that we can measure it experimentally. I think we can define it in two ways:

- *Intelligibility in the strong sense*: Language-specific metaphors are perceived to be as conceptually plausible as metaphors that are cross-linguistically shared.
- *Intelligibility in the weak sense*: Language-specific metaphors are perceived to be conceptually more plausible than novel metaphors.

The dependent variable is *conceptual plausibility* and the plausibility judgement task is phrased as “how much sense does a metaphor make” as a measure of cross-linguistic intelligibility in all of the experiments. This judgement task is relativistic, i.e. conclusions can only be drawn from the displacement of plausibility intuitions by modulation of the independent variables, which are:

- *Language proficiency of participants* (English monolinguals, German monolinguals, English-German bilinguals)
- *Language-specificity of metaphoric, idiomatic meanings* (English-specific, German-specific, cross-linguistically shared, novel)
- *Mode of presentation* (out of context, in context)

That is to say, we cannot hope that when it should be the case that participants judge L2 metaphors (metaphors specific to the language that is not their native language) to be conceptually as plausible as cross-linguistically shared metaphors that this warrants the conclusion that they are highly intelligible and cross-linguistically communicative in an absolute sense. Equally, when it should be the case that participants find L2 metaphors to be conceptually as plausible as novel metaphors this should not make us inclined to think that the metaphors are unintelligible. In both cases, it only means that, given the conditions of the experiment (the setting of the independent variables), participants may not have a reason to *express* a difference

in plausibility they perceive between the metaphors. When the conditions change, they may have a reason to do so and it is only when we see a change in their rating behaviour that we can draw conclusions about the underlying motivations. This is what I mean when I say the plausibility judgements are *relativistic* and we need to draw conclusions from *the displacement of intuitions*. At this point I cannot give any better reason other than to ask that you be patient until we look at the experimental results where, I think, these facts about the nature of the plausibility judgement task will become clear. In a very true sense, one of the central points of this thesis is what kind of *plausibility* participants are actually judging under the different experimental conditions. In the experimental results it will become clear, I think, that participants are not judging plausibility on purely conceptual grounds but that their judgements are influenced by language-specific idiomatic meanings and relevance expectations whenever these are available to them.

4.4.2 Speaker groups

Language proficiency is the first of the three independent variables. For the purpose of the experiments *monolinguality* is defined as follows: Speakers are monolingual if they are L1 native speakers and they do not speak the L2 or well below a level of fluency necessary for the metaphors in the experimental material. *Bilinguality* is defined as follows: In the experiments, almost all bilinguals are advanced L2 learners (near natives); a few are natural bilinguals where one parent or both parents are native speakers of the L2. In order to categorise participants in the experiments accordingly, language proficiency was recorded as the number of years of institutional instruction in the non-native language, the duration of time spent abroad, language proficiency of parents and frequency of L2 use. Bilinguals include advanced second language learners with an extended stay or permanent residence in the L2 country. For the experiments that follow in Chapters 7 (out-of-context metaphor presentation) and 8 (in-context metaphor presentation) this gives us the following number of participants and gender distribution (Table 4.1). The out-of-context experiments pertain to question 1 and the in-context experiments to question 2.

Table 4.1: Total number of participants (number of women) per experiment.

Participants	E mono	G mono	Biling
Out-of-context experiments	13 (10)	19 (12)	60 (34)
In-context experiments	59 (44)	20 (10)	21 (10)

Bilinguals serve as the control group. They are familiar with both the language-specific and cross-linguistically shared metaphors whereas monolinguals will only be familiar with the cross-linguistically shared metaphors and those language-specific metaphors they know from their L1. Bilinguals should therefore find the language-specific metaphors to be conceptually as plausible as the shared metaphors. If monolinguals should find L2 metaphors to be conceptually less plausible than shared metaphors then we can compare the magnitude of this diminished plausibility to the bilinguals, making them a control group. In both studies, we therefore have three control groups: monolinguals' ratings for the shared metaphors in the judgement task (strong intelligibility), their ratings for the novel metaphors (weak intelligibility), and bilinguals' ratings in the judgement task for language-specific and shared metaphors.

We want participants in the experiments to be familiar with the metaphors selected from their native language. I propose that we rely on metaphors' frequency in corpora of English and German as an indicator of familiarity. Linguistic corpora of written and spoken language represent a sample of the language that speakers have been and are exposed to. Hence, if we choose metaphors that are frequent in these corpora then we can assume that that is because they are frequent in the language that speakers are exposed to and we want all participants in the experiments that we will look at later to have had a similar chance to encounter these metaphors and if we can we might even want to ensure that this chance has been stable throughout their lifetime. Therefore it is important to know the average age of participants so that we know for which time period to check metaphor frequency. Table (4.2) gives the mean age (and standard deviation in brackets) for each of the experiments.

Table 4.2: Mean age of participants (standard deviation) per experiment.

Participants' mean age	E mono	G mono	Biling
Out-of-context experiments	19 (1.2)	41 (16.2)	39 (15.4)
In-context experiments	26 (5.7)	36 (15.3)	35 (14.1)

Overall, the average participant is 41 years old or younger. Note that the English monolinguals in the out-of-context and in-context experiments are younger and vary less in age than the German monolinguals and the bilinguals. This is because the English monolinguals are all university students whereas the German monolinguals and bilinguals include students as well as professionals.

4.4.3 Measuring cross-linguistic metaphor intelligibility

The dependent variable, the cross-linguistic intelligibility of idiomatic metaphoric expressions, is measured in three ways in this study: (1) through reading/response times, (2) through metaphor plausibility judgements, and (3) through a context creation task. The independent variables are: (1) participants' language proficiency (English monolinguals, German monolinguals, English-German bilinguals), (2) the language specificity of metaphoric idiomatic meanings (English-specific, German-specific, cross-linguistically shared, and unconventionalized-novel), and (3) the mode of presentation (out of context and in context).

The rationale of measurand 1, reading/response times: If knowledge of the idiomatic meaning acts as an inferential shortcut, native speakers, who have this knowledge, but not non-native speakers, because they lack this knowledge, should be faster at reading and judging metaphoric proverbs which require knowledge of idiomatic meanings. CMT would predict that because of the closeness of language contact between English and German and the exchange of concepts and conceptual metaphors that comes with it, when native speakers of either English or German make use of the same basic concepts and conceptual metaphors for a particular proverb, that proverb should be cross-linguistically intelligible to non-native speakers; in other words, non-native speakers should be able to infer a meaning similar to the one that native speakers have through convention (the idiomatic meaning). Non-native speakers should therefore not show slower reading/response times for proverbs they do not know from their native language, but read them as fast as proverbs that are cross-linguistically shared between English and German. RT and GS, on the other hand, would predict that if the idiomatic meaning is not purely a

matter of conception but requires knowledge of linguistic convention, then non-native speakers who lack this knowledge should be unable to use the inferential shortcut that native speakers have through their knowledge of the idiomatic meaning and non-native speakers should therefore take longer comprehending L2 language-specific proverbs than L1 and shared ones because they have to consider all *conceptually plausible* interpretations while native speakers simply default to the idiomatic meaning.

The rationale of measurand 2, metaphor plausibility judgements: Metaphors are communicative if they are intelligible and in order to be intelligible they must first be plausible. Therefore, a metaphor whose implied analogy between source and target is implausible is also not intelligible and not communicative. The analogy is plausible only if speakers feel justified making the necessary assumptions as part of pragmatic enrichment. Metaphor plausibility judgements can therefore be viewed as an index of their intelligibility and communicativeness. Because of the closeness of language contact between English and German and the concepts and conceptual metaphors that are shared between them because of this closeness, when speakers of both languages share the basic concepts and conceptual metaphors for a particular metaphoric proverb, that proverb should be as intelligible to non-native speakers as it is to native speakers and this, CMT would predict, should be true by virtue of conceptual plausibility. RT and GS, however, would predict that the proverb should not be cross-linguistically intelligible on the basis of conception alone, but only if speakers know the linguistic conventions associated with it: the idiomatic meaning, omitted context, and expectations of contextual relevance. RT and GS would thus predict that non-native speakers should not find proverbs from the other language as plausible as proverbs from their native language because their considerations of ‘plausibility’ include their knowledge of linguistic conventions (or lack thereof) *on top of* purely conceptual considerations and they would predict that considerations of linguistic convention should take precedence over purely conceptual considerations.

Meaning is well-known to be highly context-dependent. In its strongest version, RT even goes so far as to suggest that linguistic expressions may have little to no core meanings outside of context. In this study, metaphor plausibility judgements are therefore measured in and out of context. As we will see, context affects the plausibility judgements of the two control conditions. Since measurand 2 defines weak and strong intelligibility relative to the two control conditions, we need to evaluate weak and strong intelligibility as the rate of change in plausibility ratings of language-specific metaphors *as a function of* the rate of change in the plausibility ratings of the control conditions, the change being the change from out-of-context to in-context metaphor presentation. The context given to non-native speakers is intended to improve metaphor plausibility and is partly created by native speaker participants and partly by the researcher. CMT, RT, and GS agree when contextual information is relevant: it is relevant if it yields positive cognitive effects, is informative, contributes something new to the discourse, aids comprehension, or is an explanation of conceptual metaphoric mappings, i.e. if it is a substitute for the conceptual metaphors that speakers might lack. Context should make it easier to form *ad hoc* concepts. In other words, context is relevant if it is helpful during comprehension. If contextual information is relevant for comprehending a metaphor, then (all) speakers should (always) make use of it. If contextual information is irrelevant for comprehending a metaphor or conceptually conflicts with what speakers already inferred, speakers should ignore this new contextual information. With relevant contextual information, cross-linguistically unintelligible metaphors should become intelligible.

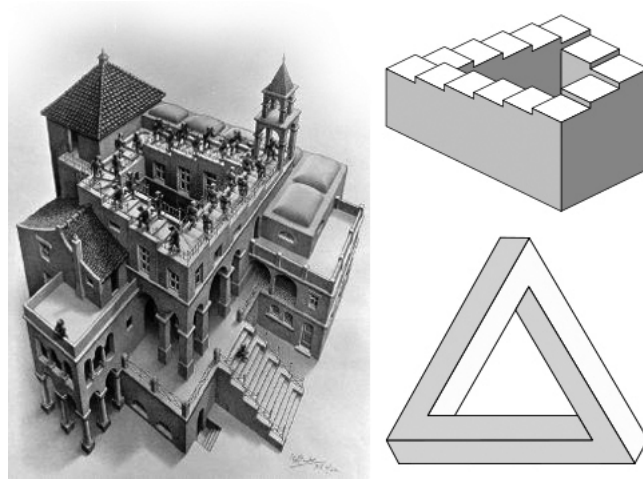


Figure 4.1: Seen on the right, an impossible triangle and an impossible staircase, originally created by Oscar Reutersvärd in 1934 and independently reinvented by Roger Penrose and his son in 1959. Inspired by Penrose's 1959 book, M.C. Escher created his famous lithograph *Ascending and Descending* in 1960 seen on the left.

A linguistic judgement consists of two parts: (1) an intuition and (2) a rationale of that intuition. I intentionally say *of that intuition* not *for that intuition* because the rationale is independent of the intuition: it is not its cause, and it may often come after the intuition. Kahneman (2002) and Kahneman, Slovic & Tversky (1982) give a very useful characterisation of this distinction. Optical illusions are an excellent example for showing that *intuition* and *rationale* are independent of each other: the infinite staircase, in Figure (4.1), is the illusion that a flight of stairs, on the one hand, seems to continue going upward, but, on the other hand, when we follow its path we find ourselves back at the foot of the staircase. It forms an infinite loop. There are many variations to this. In some, the staircase changes the direction of the footpath, but in all versions the illusion is achieved by exploiting a simpler optical illusion, the most simplest of which is the Penrose triangle (also shown in 4.1): whenever a volumetric object is projected onto a plane, one degree of freedom, the depth dimension, i.e. distance to the eye of the observer, is lost. As a result there is an ambiguity for every point in the planar projection whether that point is near or far. The infinite staircase illusion exploits this ambiguity by incrementally changing reference points within the planar projection, thus suggesting that the steps of the staircase are moving further away, only to change the point of reference later on so as to suggest that other steps of the staircase are near again. Another way to understand how the illusion is achieved is to look at overlapping surfaces. If we start at the top corner of the Penrose triangle, for instance, it seems the face to its right is the outside face of the triangle, making the opposite face, the inside face of the triangle. If we follow what we perceive as the outside face to the bottom-right corner and then move on to the third corner, what we earlier perceived as the outside face when we looked at the first corner, still looks like the outside face, but with a different orientation. This makes the new third face the inside face. But once we return to the top corner, this new inside face must change its orientation in order to plausibly be the lefthand face of the top corner. So each of the three faces seems to have two orientations,

depending on which adjacent corner we look at. We decide on the 3-dimensional orientation of the object based on two reference points, e.g. two corners of the triangle, but once we take another reference point into consideration, this third reference point suggests a new orientation that conflicts with our original assumption of the orientation.

The important insight of the Penrose triangle and staircase is that even observers who have seen the illusion many times and understand exactly how the illusion is constructed, are nevertheless *unable to not see the illusion*. Those who understand the mechanics of the illusion are often perfectly able to rationally explain its workings and can see both interpretations of the optical ambiguity, but they are nevertheless unable to not experience the illusion in the first place. This tells us three very important things about the nature of intuitions: (1) intuition precedes rationale, (2) intuition is independent of the rationale, and (3) a rational understanding of the mechanics of the intuition does not prevent us from having that intuition; it does not render the experience of the intuition mute. In experiments we assume that intuitions reveal more about the workings of the mind than the rational explanation of those intuitions and so we aim to design tasks in behavioural experiments to be such that they elicit intuitions and record people's intuitions rather than their rationale.

We are also more interested in intuitions because the rationale has a danger of being a *post hoc ergo propter hoc* rationale, i.e. people are not sure why they have a particular intuition so they make up a rationale, a rational explanation, that is a plausible explanation of that intuition even though it may be different from how they actually came to have that intuition and people think the made-up rational explanation is justified because it makes the right predictions, but without actually considering all alternative hypotheses (*confirmation bias*). Participants may not even be aware that that is what they are doing when they verbalise the rationale. This limits people's ability to assess their own inferences. In other words, people's capacity of *introspection*, the ability to analyse one's own mental states and inferential processes, is limited and often biased to confirm rational expectations. An obvious reason for this behaviour lies in our need to rationalise. Intuitions such as in the case of optical illusions show themselves to lead us astray. In our need to not "fall for the trick" we look for any rational explanation of what is "really going on" and once we have found any rational explanation we hold on to it as long as it confirms our assumptions and predictions (*confirmation bias*). Ideally in an experiment, we want to formulate a task such that we measure speakers' intuitions rather than their rationale, thereby minimising these risks, and the general assumption in psycholinguistics is that RTs tap into our intuitions more directly than judgements because we have less control over minute changes in mental processing speed than we have over rational explanations (Jegerski, 2014). I therefore use both RTs and plausibility judgements so that we may compare them.

The rationale of mesurand 3, contextual continuations: If comprehending a particular metaphor requires knowledge of its idiomatic meaning and the idiomatic meaning can be thought of as the context omitted from overt linguistic expression that would be necessary to 'spell out' the full conventional speaker meaning in a non-idiomatic way, then native speakers who know the intended full conventional meaning, should be able to make (at least part of) the omitted context explicit. Since the idiomatic meaning also includes expectations as to which contextual information would be helpful during comprehension, when speakers are asked to create sensible context for a metaphoric proverb, only native speakers but not non-native speakers should be able to meet these conventional expectations. If it is appropriate to think of idiomatic meanings as language-specific conceptual metaphors, then non-native speakers should be able

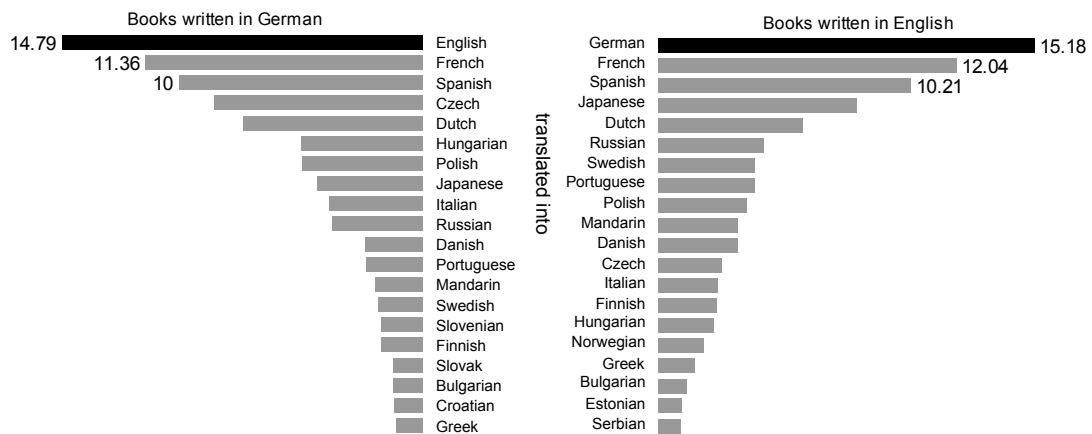


Figure 4.2: Percentage of books written in German and written in English, translated into another language since the year 1970. Data taken from the UNESCO’s Index Translationum (2013).

to make helpful context explicit on the basis of their conceptual knowledge and considerations of conceptual plausibility whenever they share the necessary basic concepts and conceptual metaphors with native speakers. CMT would propose that idiomatic meanings are language-specific conceptual metaphors and would therefore predict that non-native speakers should be able to do this for two languages where, for the proverbs in question, basic concepts and conceptual metaphors are cross-linguistically shared because of intense language contact. RT and GS, on the other hand, see the idiomatic meaning, the knowledge of omitted context, and expectations of helpful context as part of arbitrary linguistic convention and not conception, and not necessarily as motivated by Embodied Cognition. RT and GS would therefore predict that non-native speakers should fail to make context explicit which is relevant to the conventional meaning of an L2 proverb.

4.4.4 Closeness of language contact between English and German

Closeness of language contact promotes the exchange of concepts. Book translations can be understood as a good measure of the kind of intercultural exchange of concepts that we are interested in because translation is the active process of communicating that which is unfamiliar, that which has no analogue in the other language and it is therefore the act of communicating *new* concepts. Therefore we can assume that the more intercultural exchange of concepts there is between any two languages, the more the conceptual systems of their speakers align and book translations can be understood as a measure of the magnitude of that exchange. Book translations can therefore be seen as a quantitative measure of language contact and maybe even an objective one, albeit only one indicator. Other indicators would, for instance, be other forms of entertainment. At least in the direction English → German, one can say that Germany literally translates *every* film and television series it imports, with no time delay for films and maybe a one-year delay for television, facilitating a rapid cultural exchange of new concepts. Figure (4.2) shows the top 20 languages for English and German books to be translated into. The data is from the UNESCO’s Index Translationum (2013); so these are the top 20 worldwide.

Notice that these are book translations over the last 43 years which is a little more than the mean age of participants, which is around 35, and so we are particularly keen on ensuring that the requirement of closeness of language contact is met for this time frame.

As we can clearly see from Figure (4.2), English is the number one language for German literature to be translated into. This does not come as a surprise given the international status of English as a *lingua franca* in politics, commerce, and entertainment, but it may come as a surprise that German is the number one language for English literature to be translated into. Looking at this indicator of language contact we can see that English and German should have one of the most intense intercultural exchanges of concepts.

4.4.5 Language specificity of linguistic metaphors

In order to experimentally test the theoretical predictions we need to identify metaphors that are specific to a language, i.e. occur only in a particular language, either English or German in our case, and we also need to identify two other types of metaphors as controls: (1) cross-linguistically shared metaphors that occur in both English and German (*strong intelligibility*) and (2) novel metaphors that occur neither in English nor German (*weak intelligibility*). Cross-linguistically shared metaphors have a conventional linguistic form in English and German, but for language-specific metaphors we will have to translate them into the L2: German-specific will have to be translated into English and English-specific metaphors into German. We therefore need to ensure that the translated metaphors are cross-linguistically comparable in terms of their linguistic form and that their linguistic meaning differs only with regards to the language-specific, idiomatic meanings. Since the novel metaphors are new to speakers of English and German alike, we will have to ensure their cross-linguistic comparability of form and meaning as well. Table (4.3) gives examples for the four metaphor types used in both studies; the continuation examples given for the in-context experiments are the ones that correspond to the example used in the out-of-context experiments.

Table 4.3: Examples of the different metaphor types: English-specific (E), German-specific (G), cross-linguistically shared (S), and novel (N).

Metaphors	E	G	S	N
Out-of-context experiments	Talk is cheap	Sport(s) is murder	Time is money	A home is a sponge
In-context experiments	... in its exuberance	... for the non-athletic	... in the working world	... of comfort

Earlier I suggested that we use metaphors' corpus frequency as an indicator of their familiarity. To this end I used the following six English and three German corpora:

English corpora (~300 billion words worth of text):

British National Corpus (BNC)

Corpus of Contemporary American English (COCA)

Corpus of Historical American English (COHA)

Google Books UK n-grams

Google Books US n-grams

Google Books Fiction n-grams (mostly American English)

German corpora (~40 billion words worth of text):

Corpus of Contemporary German from the COSMAS-II project (COSMAS2-W)

Corpus of Historical German from the COSMAS-II project (COSMAS2-H)

Google Books German n-grams

I selected these corpora with the following criteria in mind: (1) whenever possible we want to use corpora that have been parsed and tagged for part-of-speech because they are (1a) generally more carefully parsed than non-part-of-speech tagged corpora such as n-gram corpora (I will discuss the multitude of spelling mistakes in n-gram corpora below), and (1b) when we look at the frequency of the words that the metaphors are composed of, at the end of Section (5.4), we can use the part-of-speech tagging to distinguish between, for instance, the noun *Reden* (German *speeches*, noun, plural; or the verb *to talk*, *to speak* that has been capitalised because it occurs at the beginning of a sentence; or the verb *to talk*, *to speak* that occurs in the middle of a sentence but has been capitalised orthographically to indicate that it has been nominalised, cf. *the speaking* in English) from the verb *reden* (German *to talk*, *to speak*, verb, infinitive) for the translation of the word *talk* in the English metaphor *talk is cheap*. (2) For each language we want at least two corpora so that we can ensure the reliability of the frequency counts we get. (3) In order to ensure that participants in the experiments have had a similar and stable chance to familiarise themselves in their lifetime with the metaphors we expect them to know, we need to conduct both a synchronic and diachronic frequency analysis and we therefore need both synchronic and diachronic corpora for each language and dialect. (4) We want the largest corpora available because the bigger a corpus is in size the more reliable the frequency counts are (all other things being equal, such as part-of-speech tagging versus non-tagging, of course). (5) As idiomatic linguistic expressions we should not only expect metaphors to vary across languages but also across dialects: for English, I have selected corpora representative of British versus American English. For German, there are either no dialect-specific corpora available or they are insufficiently large or they are of reasonable size but their texts are not rich with the proverbial metaphors we are interested in (for instance, because they are composed of regional newspapers) or they are not contemporary dialect-corpora, i.e. they are much older than the average age of participants. At the same time, by restricting ourselves to highly frequent language-specific and cross-linguistically shared metaphors we should be selecting those metaphors that are common across dialects and many of these dialect-specific issues should be less of a concern to us. Note that I have selected at least one synchronic and one diachronic corpus per dialect and language to this end.

For English, the smallest corpus is the British National Corpus (BNC) with 96.2 million words (University of Oxford, 2007). This corpus is representative of British English and is a synchronic corpus that is constructed from texts and speech collected from the late 20th century. The corpus project is a collaboration of Oxford University Press, Longman, W. & R. Chambers, the University of Oxford, Lancaster University, and the British Library. After its initial release in 1994 and revision for the second edition in 2001, the BNC is currently in its third edition released in 2007. In an attempt to ensure genre balance samples are taken from a variety of media. In this spirit the 90% of the corpus that are written samples are taken from regional and national newspapers, scientific journals and periodicals, fiction and

non-fiction books, both published and unpublished material such as leaflets, brochures, letters, student essays from different academic fields and levels of study, speeches, and scripts. The remaining 10% of the corpus are samples from spoken language that have been orthographically transcribed and are natural conversations produced by volunteers of various ages, social classes and from different parts of the UK. At the same time these conversations were produced in different contexts, including formal business or government meetings and events, conversations on radio shows and phone conversations.

The other British English corpus among the six English corpora is the Google Books UK n-gram corpus, a diachronic corpus with 50.8 billion words (Davies, 2011c; Michel*, Shen, Aiden, Veres, Gray, The Google Books Team, Pickett, Hoiberg, Clancy, Norvig, Orwant, Pinker, Nowak & Aiden*, 2011; Google, 2012). The following applies to all of the Google n-gram corpora. The corpus material is less balanced than in the BNC because it includes a large sample of books from Google Books but not from other written media of language such as newspapers. The only restriction is that the UK n-gram corpus is sampled from books published in the UK and the US corpus from books published in the US. The Fiction n-gram corpus, according to Google's own statement on their website (Google, 2012; Davies, 2011b), consists of fiction books published both in the US and UK, but the quantity of US fiction books is much greater. For that reason I will assume in this thesis that the Fiction n-gram corpus is primarily representative of American English. The Google Books n-gram corpora are currently in their second version published in 2012. All the data in this dissertation is based on the n-grams from the first version released in 2009 because the corpus analysis was conducted in the summer and fall of 2012 before the release of version two. There are some notable differences, many of which I will point out as flaws in the discussion of the corpus analysis. In version one, because the books were scanned using 20th century OCR (Optical Character Recognition, i.e. text recognition software) elongated medial *f* in Middle English texts was always parsed as lowercase *f*. This has now been corrected in version two. Also, version 1 does not contain any part-of-speech tags which may lead to structural and syntactic ambiguity in cases where the limitation of context prevents manual disambiguation (the length of n-grams is limited to strings of five words and punctuation such as commas and brackets as well as numerals count as words too). Version 2 now contains part-of-speech tags for nouns, verbs, adjectives, adverbs, pronouns, determiners, adpositions (prepositions and postpositions), numerals, conjunctions, and particles as well as three structural tags for the root of a parse tree, the start and the end of a sentence. In addition to these tags, version two allows the user to search for syntactic dependencies within the parse trees using lexical items, tags, or a combination of both and a set of operators to express the syntactic relationship. Since these tags and operators were not available to me at the time I present an iterative filtering algorithm that filters out parsing errors and iteratively disambiguates and refines the search criteria to get precise frequency counts and I will give a more detailed description of this algorithm later on.

As a synchronic corpus for American English I used the Corpus of Contemporary American English (COCA) which contains 464.3 million words worth of text annotated with part-of-speech tags (Davies, 2008). It was created by Mark Davies and is curated by the Brigham Young University with written and spoken language from 1990 to 2012. For each year of texts, the corpus is evenly divided between the five genres of spoken, fiction, popular magazines, newspapers, and academic journals. Spoken language is orthographic transcriptions of natural unscripted conversations from 150 different TV and radio programmes. Fiction is short stories

and plays, from literary magazines, children’s magazines, popular magazines, first chapters of first edition books published between 1990 and 2012, and movie scripts. Popular magazines comprises texts from nearly 100 different magazines with a mix from different domains such as news, health, home and gardening, women, financial, religion, and sports. Newspapers are ten regional and national newspapers. There is a good mix between different sections of the newspaper such as local news, opinion, sports, and financial. Academic journals are nearly 100 different peer-reviewed journals selected so as to cover the entire Library of Congress classification system (e.g. a certain percentage from B (philosophy, psychology, religion), D (world history), K (education), T (technology), etc.).

As diachronic corpora for American English I used the Corpus of Historical American English (COHA) which is 407.6 million words in size and contains part-of-speech tags (Davies, 2010) and the two Google Books n-gram corpora of fiction books and books published in the US of 90.7 billion and 157 billion words respectively (Davies, 2011a,c; Michel* et al., 2011). The COHA was created by Mark Davies and is curated by the Brigham Young University. It uses the same part-of-speech annotation system as the COCA. The COHA contains texts from 1810 to 2009. The corpus is divided into four genres: fiction, magazines, newspapers, non-fiction. Fiction contains US texts written between 1810 and 1930 from Project Gutenberg, the Making of America (1810-1900), scanned books (1930-1990), scripts for movies and plays from COCA (1990-2010) and is balanced for the sub-genres prose, poetry, drama, etc. Magazines comprises texts from the Making of America (1810-1900), scanned books and PDF files (1900-1990), and magazines from COCA (1990-2010) and is balanced for domains such as health, home and gardening, financial, etc. Newspapers are scans of at least five newspapers (1850-1980) and newspapers from COCA (1990-2010) with balanced sub-sections. Texts in non-fiction are taken from Project Gutenberg (1810-1900), www.archive.org (1810-1900), scanned books (1900-1990), and COCA (1990-2010) and in each decade, the text selection is balanced across the Library of Congress classification system.

The smallest of the German corpora is the historical corpus denoted as archive HIST from the COSMAS2 corpus collection with only 4.3 million words (Institut für Deutsche Sprache, 2013). Metaphors are over-represented in COSMAS-H (the historical archive of the COSMAS2 corpus project) and this is most likely because a large part of the 4.3 million words in that archive are taken from plays written by Goethe and Schiller, both rich in metaphors. So not only is the historical German corpus quite small, it is also highly unbalanced in terms of genre representation. All the other corpora also include more prose and news reports as well as academic papers—all genres that contain less metaphors and thus counterbalance the metaphor frequencies. When we exclude the data from COSMAS-H, the frequency counts of the language-specific and cross-linguistically shared metaphors are quite comparable between the English and German corpora (note, they are *included* in Table 5.2). Luckily we have another diachronic corpus for German, the Google Books German n-gram corpus with 37 billion words (Michel* et al., 2011; Google, 2012). As a synchronic corpus we have the corpus of written language, denoted as archive W, from the COSMAS2 project with 2.8 billion words. The COSMAS2 corpora are annotated for part-of-speech and syntactic structure using a tag syntax similar to the BNC. COSMAS2 is a project of the Institut für Deutsche Sprache (Institute for the German Language) and is curated by the University of Mannheim. It was first released in 1992 and more texts continue to be added every year. It is a collaboration with regional and national newspapers and publishing houses in an attempt to provide a representative variety

of balanced text genres. It includes written texts from regional and national newspapers, journals and periodicals, biographies, manuscripts kept by the Gesellschaft für Deutsche Sprache (the Society for the German Language), fiction and non-fiction books including manuals, the collected works of Goethe, Marx, Engels, Mann and the brothers Grimm, orthographically transcribed interviews and speeches, political rhetoric and propaganda. The majority of the material comes from the Republic of Germany, but also includes some Austrian newspapers.

Corpora with part-of-speech and syntactic annotation are qualitatively more reliable than those without annotation. Concretely this means that we should be more critical about the frequency counts we acquire from the Google corpora (in their first version since the second version features part-of-speech tags and basic syntactic constituency), but there are two features to the Google corpora that I think make up for this: (1) they are much larger than the annotated corpora. Recall that the largest annotated corpora do not exceed half a billion words, while the “smallest” Google corpus is 37 billion words and the largest one is over 157 billion words in size. In other words, the Google corpora are between 80 and 300 times larger than the tagged corpora.

The second feature of the Google corpora that I think makes up for the lack of annotation is (2) that the metaphors we are investigating, the cross-linguistically shared and language-specific metaphors alike, are all quite fixed idiomatic expressions, some of which originate in proverbs, others are famous quotations, e.g. from Shakespeare, that have become idioms. Because they are so formally fixed part-of-speech tagging and syntactic annotation are not always necessary and in cases where multiple versions or phrasings of a metaphor exist, as in the case of *an Englishman’s home is his castle* (*a man’s home is his castle*, *my home is my castle*, etc.), we can simply search for all known instances and accumulate the frequency counts. And in cases where we might not be aware of all possible phrasings, even the simplistically parsed data of the Google corpora in version one, where orthographic words are separated from one another and from punctuation and numerals, but nothing more; even in this case is the parsed string of tokens (here meaning the sum of all words, punctuation marks, and numerals) enough to perform searches which allow intermitted material so as to anticipate possible alternative phrasings. I will detail my design of the search engine I wrote to look for the metaphors in the Google corpora, but before we get to these workarounds (necessitated only by the shortcoming of the lack of part-of-speech tagging), let us outline the formal criteria for selecting the metaphors in the first place.

I used the online search engine developed by Mark Davies (Davies, 2004, 2008, 2010, 2011c,a,b) to look up metaphor frequencies in the English corpora, including the British National Corpus (University of Oxford, 2007), COCA, COHA, and the Google n-gram corpora (Google, 2012; Michel* et al., 2011). The German corpora of historical and contemporary German from the COSMAS II corpus collection uses the online search engine developed by the Institute für Deutsche Sprache (2013). There was no search engine for the German Google books n-gram corpus at the time and the Google proprietary search engine Ngram Viewer does not give exact token counts but rounded percentages. I therefore developed my own search engine in Python (van Rossum & Python Software Foundation, 1991) which uses the same filters as the BYU and COSMAS II Interfaces. See table A.3 in appendix A for a complete list the metaphor frequencies and table A.2 for the frequency of the words in the metaphors. For the frequency distribution of the entire American English, British English, and German vocabularies I downloaded the original n-gram datasets from Google (2012) and wrote my own search

engine in Python (van Rossum & Python Software Foundation, 1991) to perform “number of wordforms per frequency band” look-ups. This frequency distribution will become relevant when we compare the frequency of language-specific and cross-linguistically shared metaphors cross-linguistically and will be explained in more detail later on in this chapter.

Not only is it good practice to have several corpora for each language so that we can compare the consistency of frequency counts by comparing between corpora, but it also allows us to spot potential lexicalisation differences between British English and American English and allows us to compare synchronic and diachronic frequency developments. Primarily we are interested in ensuring that metaphors fall clearly into one of the four categories, but we are also interested in ensuring that participants are likely to be familiar with them and if we know that a particular metaphor has been, say, an English-specific metaphor for the last 150 years—meaning it never occurred in German over the same period of time—then it is reasonable to assume that the categorical differences we hope for in preparing our experimental material will be strong enough to yield different intuitions in participants.

The corpora are going to be a means to justify the classification of the metaphors into cross-linguistically shared, English-specific, German-specific, and novel metaphors; but for the experiments we want all metaphors to be formally similar so that it will be more likely that they will have the same processing complexity. We also want them to be semantically similar, but since we differentiate between language-specific and cross-linguistically shared metaphors at least as a theoretic working hypothesis we also presuppose that they will be semantically different with respect to the metaphors’ idiomatic meanings. By calling some metaphors English-specific we presuppose that that only speakers of English will be aware of the metaphors’ idiomatic meanings because idiomatic meanings are arbitrary (not inferable through common sense and analogy alone) and must thus be acquired as part of the language. So for instance, speakers of another language, in our case German, who do not speak English (or not reasonably well enough to have acquired the metaphors’ idiomatic meanings), these speakers will have to rely on other means to infer a meaning for the English-specific metaphors, e.g. by singling out relevant information from their world knowledge by use of common sense, drawing analogies, and assessment of the plausibility of potential metaphor interpretations. Similarly, for German-specific metaphors German native speakers would have an advantage because of their knowledge of the idiomatic meanings, while speakers of English who do not speak German (well enough) would have a disadvantage. And for cross-linguistically shared metaphors speakers of English and German should be equally able to incorporate the metaphors’ idiomatic meanings during the inferential comprehension process.

The availability of idiomatic meanings during the inferential process is thus one of the experimental conditions we wish to control through comparison of participants of different language proficiencies and since the availability of idiomatic meanings is a semantic feature, if we wish to manipulate it experimentally, we must ensure that all other semantic aspects, to all intents and purposes in the experiments, remain the same. The other semantic aspects are: (1) lexical semantics, specifically how similar are the meanings of the individual words in the translations of the metaphors between English and German in terms of the number and kind of senses, intentions, and extensions; and (2) compositional semantics, i.e. the way in which the relationships that result from combining the individual lexical items into larger units contribute to the meaning of the whole. Since this research focusses on the semantic sources of information that interlocutors make use of during the comprehension of metaphors, it is

reasonable to minimize the metaphors' syntactic complexity and variation.

We now have the terminology to specify what, formally speaking, the simplest metaphor format looks like: first, a noun phrase that denotes the target concept, followed by the copula, followed by the second noun phrase which denotes the vehicle concept. Notice how the syntax of the sentence mirrors the semantics here: because of the copula the second noun phrase is predicated over the first noun phrase which is exactly the directionality of the implied semantic analogy. Apart from noun phrases, adjective phrases can also occur in this predicative position following the copula and so it is reasonable to assume that the adjective *blind* in the sentence *love is blind* is predicated over the target concept LOVE in the same way that a nominal vehicle would be, and so we also allow metaphors where the vehicle is an adjective phrase to be part of our investigation, while still upholding our objective to keep the form as minimalistic as possible in order to focus on the semantics and pragmatics of the inferential process. The way that syntax and semantics are mirror images of each other in the case of this metaphor format also means that we know that compositional semantics will always be along the directionality implied by the metaphoric analogy and the syntax.

4.4.6 Context and fillers

In most psycholinguistic experiments (e.g. Keysar & Glucksberg, 2000; Gibbs, 1992) context precedes the expressions under investigation. In the experiments of Chapter 8 of this thesis, context follows rather than precedes the metaphors under test. Let me motivate this decision: yes, it could be argued that this means the inferred metaphor meanings are the result of re-analysis rather than a reflection of speakers' initial understanding of the metaphors. A meaning that is the result of re-analysis would necessarily be different from an original, initial meaning inferred. The more time speakers are given during the experimental tasks, the more time and opportunity they have to change their mind on the interpretation of a metaphor. We saw that monolingual speakers would have the concepts necessary to comprehend the L2 metaphors in away that native speakers would, and given enough time the chance that they could surely use considerations of plausibility and relevance to arrive at a more satisfactory metaphor meaning than an initial one, which, from a critical standpoint, supposedly must have been inadequate as it resulted in speakers forming such low plausibility judgements.

But we could also choose to interpret the results, including the out-of-context ones, differently: the low plausibility judgements might not be the result of speakers not having enough time to consider all possible metaphor interpretations. Rather, even when monolingual speakers considered all possible interpretations "they could think of," the most plausible meaning they arrived at inferentially was still suboptimal, less plausible, than for L1 and cross-linguistically shared metaphors. So not constraining monolinguals' time to make a plausibility judgement should actually work in favour of CMT because the more time speakers have, the more conceptually plausible alternatives they can consider. This way of reasoning applies to both the out-of-context and the in-context results. So the way that context is operationalised may not be problematic because it actually makes it easier to confirm CMT, which is however not what the results suggest.

Our main focus in this thesis is on the out-of-context results anyway. Speakers in-context rating behaviour merely shows more clearly that monolinguals behave quite differently from bilinguals, a behavioural difference which I suggested is caused by monolinguals' ignorance of the

non-conceptual aspects of L2 metaphors' idiomatic meanings. On this view, why monolinguals arrive at a metaphor meaning that is suboptimal compared to L1 and cross-linguistically shared metaphors is explained by the information, which, I argue, must be contained in the non-conceptual aspects of metaphors' language-specific idiomatic meanings; information which for the L2 metaphors, so the results seem to suggest, is available to bilinguals but not monolingual speakers.

Thus far I have introduced four metaphor types to be included in the experiments, but you may wonder why there are no fillers. It could be argued that the novel metaphors are a form of fillers. Then again, as we will see from the results in Chapter 7 and 8, cross-linguistically shared metaphors always receive the fastest reading/response times and the highest plausibility judgements, on average. Novel metaphors always have slower reading/response times and the slowest plausibility judgements. The metaphors that we are interested in, those with language-specific idiomatic meanings which do not occur in the other language, can then be analysed *relative* to the scale which is naturally provided by the deviation of reading/response times and plausibility judgements between cross-linguistically shared and novel metaphors.

As experimental researchers we assume that fillers will receive responses which are neutral, i.e. unaffected by the experimental manipulations. Thus the assumption is that fillers will receive baseline ratings. In the case of metaphor comprehension, this neutrality of fillers seems implausible. If the fillers are metaphors, how would we characterise them as "neutral" relative to the test metaphors? If the fillers are non-metaphoric, are they even adequate baseline responses against which to compare the responses of the test metaphors? Instead, I propose we take the deviation in responses between cross-linguistically shared and novel metaphors as a natural scale relative to which we can analyse the responses of the language-specific metaphors.

As we will see, the metaphor type whose responses remain most constant across all speaker groups and out-of-context/in-context conditions are the cross-linguistically shared metaphors. Their responses are therefore a reliable baseline and potentially not worse than fillers. In some sense, the cross-linguistically shared metaphors are fillers because they are (a) metaphors and (b) we would not expect their responses to be affected by the experimental manipulations. Similarly, it could be argued that the novel metaphors are fillers because the features of their linguistic form and Fregean senses are subjected to the same requirements imposed on the language-specific metaphors, yet they are different because they have no conventionally established idiomatic meanings. It could thus be argued that the cross-linguistically shared and novel metaphors represent two types of fillers: the cross-linguistically shared metaphors are fillers with language-specific idiomatic meanings that are unaffected by the experimental manipulations and should thus receive similar ratings by all speaker groups, and the novel metaphors are fillers for which no speaker group can make use of language-specific idiomatic meanings during the comprehension process.

4.5 Summary

In this chapter we have looked at the two central questions of this research. When empirical research in linguistics and cognitive science into the mechanisms underlying metaphor has done so from a cross-linguistic perspective, it has done so mostly by comparing languages with little or no linguistic contact so as to find conceptual differences between their speakers. I, however, think we should critically question the validity of the assumption that speakers of languages in

close contact will use similar concepts in comprehending the same metaphors. The first central question of this research therefore is: *Are metaphors from one language intelligible to speakers of another language if those two languages are in close linguistic contact and thus can be assumed to share similar or the same concepts?* It is well known that the context of an utterance has a great influence on the speaker meaning that is inferred. At the same time we can assume that context could supply information useful for comprehending unfamiliar metaphors. Therefore, if in pursuit of question 1 we find that some metaphors are unfamiliar to certain speakers and those metaphors are thus cross-linguistically less intelligible to the same speakers than metaphors they are familiar with, then we must ask ourselves if and how that picture would change if the same speakers were making their plausibility judgements under the influence of contextual information. I thus proposed the second central question as: *If context improves the intelligibility of metaphors across languages, do bilinguals make use of context in the same way that monolinguals do?*

I then suggested that we can define the cross-linguistic intelligibility of metaphors in two ways: a strong and a weak definition. *Intelligibility in the strong sense:* Language-specific metaphors are perceived to be as conceptually plausible as metaphors that are cross-linguistically shared. *Intelligibility in the weak sense:* Language-specific metaphors are perceived to be conceptually more plausible than novel metaphors.

Since we are only comparing metaphor comprehension in two languages, English and German, we can define monolinguality and bilinguality in a narrow sense: English monolinguals are native speakers of English (with at least one parent as a native speaker and the non-native parent is not a speaker of German) and do not speak German. German monolinguals are native speakers of German (with the same restrictions on the languages spoken by their parents) who do not speak English (neither regularly, conversationally, nor for recreation). English-German bilinguals are those who speak both English and German, either because they are native speakers of German who are advanced L2 English speakers or because they are native English speakers who are advanced L2 German speakers or because of their parents where one speaks English and the other German.

That English and German have, in no small part because of their linguistic history, closer contacts with each other than with any other languages. I showed that this closeness can be quantified by looking for instance at book translations. The cultural practice of translation, especially of literary works rich with metaphors, can be understood as the act of communicating those concepts and conceptual mappings into another languages that are unfamiliar. Translation therefore reduces conceptual differences which in turn allows us to validate one of the basic assumptions of our investigation that English and German are in close linguistic contact and their speakers should have few to no differences in the basic conceptual mappings that CMT claims are the embodied circuitry on which more complex metaphors such as metaphoric proverbs are computed.

Chapter 5

Material: cross-linguistic comparability of metaphors' linguistic form

5.1 Introduction

Our main focus is on conceptual cross-linguistic metaphor similarity as a prerequisite in looking for potentially impaired cross-linguistic communicability of L2 metaphors. Therefore, what we want to manipulate experimentally is speakers' knowledge of metaphors' language-specific idiomatic meanings, which we will do by comparing monolingual native speakers' comprehension of language-specific metaphors to their comprehension of non-language-specific metaphors. We therefore want to minimise cross-linguistic differences in metaphors' other, non-conceptual features: for instance, since we intend to measure reading/response/reaction times, we want to ensure that metaphors' length does not affect the measurement in a way that conflates with the language specificity of metaphors. And we want to avoid suprasegmental features such as rhyme which might affect plausibility judgements in a way that conflates with metaphors' language specificity. If we can make it so that cross-linguistic variation in metaphors' non-conceptual aspects is minimised and equally distributed across all metaphor types and for the English and German versions of the materials, then we can interpret the experimental results "all other things being equal." Keeping metaphors' non-conceptual aspects constant across experimental manipulations is therefore a prerequisite to properly account for conceptual deficits in the material which might be unavoidable for practical reasons. This chapter therefore focuses on metaphors' non-conceptual aspects and that they are constant across experimental manipulations before we turn our attention to metaphors' conceptual cross-linguistic similarity in the next chapter.

In the previous chapter I defined the language-specificity of metaphors. This chapter describes how metaphors were selected for the experiments based on this definition and categorised into (1) cross-linguistically shared metaphors, (2) language-specific English metaphors, (3) language-specific German metaphors, and (4) unfamiliar novel metaphors. Section (5) describes this categorisation through the use of large English (~300 billion words) and German (~40 billion words) text corpora (5.2). For the largest German corpus this required me to

create my own search engine as the only one available at the time did not provide precise frequency counts (5.3). The creation of the search engine is described in this Section, followed by an analysis of the frequency of the language-specific, non-specific, and novel metaphors in all English and German corpora (5.4). In this corpus analysis, I pay particular attention to the question of how to *define* and *measure* the cross-linguistic comparability of corpus frequencies and to this end propose to use a method that converts the raw frequency counts to cumulative probabilities, akin to z-scores which are commonly used in empirical work to compare experimental results (5.4). I then apply this method to the analysis of the metaphor frequencies as well as the frequency of the words that make up the metaphors and conclude that they both are cross-linguistically comparable.

In Section (5.5), I present a diachronic analysis of the corpus frequencies within the same English and German corpora. We want to ensure that English and German participants in the experiments, whose mean age is around 30 years, have had an equal and stable chance to be exposed to the language-specific and non-specific metaphors and that they are unfamiliar with the novel metaphors. We therefore want to ensure cross-linguistically comparable and stable corpus frequencies over time, at least for the same period of years. I show that for all metaphors their frequencies are either stable or have been increasing and for many of them this can be traced back reliably much more than 30 years, often up to 200 years.

We want participants to be unfamiliar with the novel metaphors. I show in the corpus analysis how all novel metaphors used in the experiments have zero frequency. However, the chance to find highly infrequent items in corpora increases with corpus size. The English (~300 billion words) and German (~40 billion) corpora are large, but how can we be sure that we would not find any of the novel metaphors if only the corpora were bigger (which would mean these “novel” metaphors would have to be reclassified as either language-specific or shared between English and German)? In order to give an answer we can assess the *measurement accuracy* of the largest corpora in our investigation. In Section (5.6), I show how to estimate their measurement accuracy by measuring the variation in punctuation marks which, unlike most other linguistic items, vary little in frequency while being highly frequent themselves. I show that the measurement accuracies of the largest corpora can be estimated in this way for the last 200 years which reinforces the reliability of both the synchronic and diachronic corpus analyses.

In Section (5.8), we look at grammatical differences between the English and German translation of the material that affect the length of metaphors. I give four measures of their length: the number of orthographic characters, the number of syllables, the number of morphemes, and the number of words, especially with respect to mismatches in the mass/count-noun distinction and the definiteness of noun phrases between the English and German version of the material. It turns out that metaphor length is more similar than dissimilar and that the differences that we do find are highly systematic grammatical variations with predictable and known effects.

I make a conscious effort throughout this chapter and the next to not only argue that the different aspects of linguistic form, familiarity, and meaning are cross-linguistically comparable, but to also suggest ways to quantify the *degree* of cross-linguistic comparability. This allows me to summarise these quantitative measures in Table (6.3) at the end of the next chapter. This table provides measures of the cross-linguistic dissimilarity for each of the metaphor types used in the experiments as well as an analysis of variance among the metaphor types for each type of difference so that we can return to this table once we look at the experimental results

and can gauge if any of the aspects of linguistic form, familiarity, or meaning could be the source of rating differences in the plausibility judgement task. We can also see how the cross-linguistic dissimilarities of particular aspects of linguistic form, familiarity, and meaning might accumulate, if rating differences should be motivated by a set of these aspects rather than one of them in particular.

5.2 Metaphor selection process

There are thousands of metaphoric proverbs. In order to find metaphors suitable for the experiments I devised the following selection process:

1. Start from dictionaries of English and German proverbs and sayings
2. Select those that meet the following formal restrictions:
 - 2.1 They are of the form **noun A is noun B** or **noun is adjective** where both source/vehicle and target/tenor concept of the metaphor are linguistically overt and connected syntactically by the copula form of *be* which can be considered semantically vacuous.
 - 2.2 Individual words in a metaphor as well as the metaphor as a whole need to be comparable between the respective languages in terms of their familiarity and semantics.
3. Categorise according to the following criteria:
 - 3.1 For a metaphor to be classified as language-specific English metaphor, it must only occur in English corpora (~300 billion words) and never in any German corpora (~40 billion words).
 - 3.2 For a metaphor to be classified as language-specific German metaphor, it must only occur in German corpora (~40 billion words) and never in any English corpora (~300 billion words).
 - 3.3 For a metaphor to be classified as cross-linguistically shared, it must occur in both English and German corpora (~340 billion words).
 - 3.4 For a metaphor to be classified as novel, it must not occur in any of the corpora (~340 billion words).

This gives us the following number of metaphors per type (Table 5.1):

Table 5.1: Number of the different metaphor types: English-specific (E), German-specific (G), cross-linguistically shared (S), and novel (N) as well as the total number of metaphors in each study (Σ).

Number of metaphors	E	G	S	N	Σ
Out-of-context experiments	5	12	13	11	41
In-context experiment	12	12	14	12	50

We need to translate the language-specific and novel metaphors. The cross-linguistically shared metaphors have well-established phrasings in English and German. Hence there is no need

to change them or to come up with new translations. For the language-specific and novel metaphors, however, there are no translations in the other language. Of course in order to check that the language-specific metaphors do not occur in the other language, we need to check possible alternative translations. This naturally leads us to probe whether the translations we end up with are the best-possible translations. We can define best-possible translation here as a translation where the form and meaning of a metaphor are arguably the same (for the intents and purposes of this research *explanandum*).

5.3 Corpus frequency as an indicator of speakers' familiarity with metaphors' linguistic form

We can use corpus frequency as an indicator of familiarity. We want to ensure that participants are likely to be familiar with the metaphors (L1 and shared in the case of monolinguals and L1, L2, and shared metaphors in the case of bilinguals), so we want to pick metaphors that are frequent, that are as frequent in the L1 as they are in the L2, and we want their frequencies to be stable over time—at least for the last 30 years, the mean age of participants. We want the same cross-linguistic comparability for the words in the metaphors, too. Together, these frequencies allow us to assess the formal cross-linguistic comparability.

Out of the nine corpora five are tagged for part-of-speech (the BNC, COCA, COHA, COSMAS2-W, COSMAS2-H) while the other four Google corpora (UK, US, Fiction, German) do not have any annotation in version 1; but it is also important to note that three different search engines, i.e. interfaces, were used to perform frequency look ups. We want to ensure that the frequency counts are comparable across corpora and to begin to objectively assess that, we need to make sure the search engines' algorithms work the same way (or reasonably similar). The first interface was created by Mark Davies of Brigham Young University and was originally designed as an advanced online search interface for the BNC (hence abbreviated as BNC-BYU), but has since been also become the primary interface for COCA, COHA, and the English Google n-gram corpora UK, US, and Fiction. Key features of this search engine that are important in comparison to the other two search engines are: (1) if a search query is entered in lowercase letters, e.g. time is money, the interface automatically searches for all combinations of uppercase and lowercase, all caps, etc. and accumulates these individual frequencies. All-capital letter strings might appear in headings. (2) For the BNC, COCA, and COHA, which are divided into different genres and subdomains, the search engine allows instantly comparing frequencies for a query across these genres. (3) A minimum frequency can be defined which helps reduce the noise in low-frequency items. (4) The query syntax allows a flexible combination of text strings, i.e. lexical items, and part-of-speech tags and syntactic labels which allows us to anticipate intervening material, e.g. adjectives in noun phrases or variations of *an Englishman's home is his castle* (*my home is my castle*, *a home is a castle*). (5) Except for the BNC and COCA, which are synchronic corpora of contemporary British and American English, the interface allows us to search a specific time interval (for us this is primarily the time interval of the last 200 years) and to define if frequency counts should be summed per year or decade.

The second interface was designed by Cyril Belica in 1993, originally as a pure offline Windows-based search engine, and is the interface used for all the corpora in the COSMAS2 project. Since 2005, the search engine also became accessible through an online interface and

one year later in 2006 was extended by a script language to allow more flexibility in search queries. As a server application with local front-end, the COSMAS2 project is now curated by the University of Mannheim and is maintained and constantly extended by Dipl.-Inf. Eric Seubert (lead), Lic. phil. nat. Franck Bodmer Mory, Dipl.-Inf. Peter Harders, and Dipl.-Inf. Helge Stallkamp. COSMAS2 comprises 106 corpora that are sorted into 10 archives according to the annotation style of specific corpora: of main interest for us here are (1) the archive of written corpora, denoted as COSMAS2-W(ritten) which contains text written between the 18th century and present day, 3.9 billion wordforms of which 2.8 billion are publicly accessible through the online interface (the other 1.1 billion are only accessible to the linguists at the University of Mannheim); (2) the archive of historical corpora, labelled COSMAS2-H(istorical) which comprises texts originating from the second half of the 17th century to 1962 (this overlaps partially with COSMAS2-W containing some of the same texts of the 18th century onwards) of 65 million wordforms in size of which only 4.3 million are available to the public via the online interface. The other archives are either automatically annotated corpora (three archives of corpora that use either the (3) CONNECOR-, (4) TreeTagger- or (5) MECOLB-morphosyntactic labels) or corpora that contain texts of a specific genre to study a particular language style or historical period ((6) Linguistic origins of the 20th Century, (7) corpora of the Society of the German Language, (8) Phases of the Reunification of 1989, (9) written corpora that have been excluded from COSMAS2-W because of older tag syntax, and (10) Newspaper collection of the Süddeutsche Zeitung).

Key features of the COSMAS2 search engine are: (1) it checks all possible combinations of uppercase, lowercase, and all-capitals automatically. (2) Search results can be narrowed down to text genres. (3) A minimum frequency threshold cannot be set which is an important difference to the BYU interface, but at the end of the day it means that the COSMAS2 engine includes even tokens of a search query found only once in the entire corpus. (4) Unfortunately only 40% of COSMAS2-W is part-of-speech tagged. I therefore performed the metaphor search using lexical items only. However the interface has another feature that allows us to do what the mix of lexical items and tags in the BYU engine achieves, namely that we can specify how many words can come between each pair of neighbouring words, e.g. *love is blind* has two pairs of neighbouring words: *love is* and *is blind* and for each of those the COSMAS2 search engine allows us to specify whether we only want to count text strings where the words in a pair follow immediately one after the other, whether we allow one, two, three, etc. words to come between them, and whether the words in a pair are connected by logical AND or logical OR. If we chose AND, only if all words in the metaphor *love is blind* occur together (with potentially other words in between), does the search engine count this as one token, but not if only *love is* occur together or only *is blind*. If we chose OR instead, the search engine would count tokens where at least *love is* and *is blind* occur even if the third word is not present in those cases. This workaround is not ideal because we cannot define what part-of-speech an intervening word should be, but it is still a reasonable approximation to the capabilities of the BYU interface. (5) The interface always allows us to limit the time interval and offers the same flexibility as the BYU interface as to whether we want frequency counts to be summed by year or by decade.

For the German Google n-gram corpus the only search engine available at the time was the Google Ngram Viewer. The Ngram Viewer has a number of deficits: It is purely a tool for visually inspecting frequency ratios online. Once we have performed a search the Viewer plots the frequency ratio of one of our German metaphors, e.g. *Zeit ist Geld* (*time is money*),

meaning the frequency is expressed as the number of tokens of that metaphor divided by the number of words in the corpus for that year as a percentage. These percentages end up being incredibly tiny which can make evaluating and interpreting relative changes in frequency counter-intuitive; but worse is that the percentages are rounded so that even if we could read off the exact percentage for one year, we would not be able to calculate the exact number of tokens. This impressionistic plot cannot be downloaded. At least in version 1, the online tool does not automatically check multiple lowercase, uppercase, first-letter-capital, and all-capitals combinations. Instead we get individual frequency ratio curves. Version 2, released in 2012 now features operators that allow us to sum these individual curves to one overall curve, but we still have to search for all case-sensitive versions manually.

In texts written before the 19th century, there are systematic parsing errors produced by the OCR (Optical Character Recognition, i.e. a text recognition algorithm) where elongated *f* is always parsed as lowercase *f*. This is easy enough to identify in mis-parses of *Engliffh* as *Engliffh*, but if we take *cafe* and *cafe* where *case* would also have been spelled with the elongated *f* because it occurs in the middle of the word, there is no way of telling how many of the *cafes* are actually *cafes* in texts before 1800. Since version 1, released in 2009, also had no part-of-speech tagging I could also not use context to disambiguate those cases of *cafes/cafes* because without part-of-speech tags we would have to imagine at least the most common lexical items that would occur in the immediate context of *cafe*, but not *case* and vice versa. This quickly turns into a logistical nightmare with too many unknown variables, which is exactly why we would want to use part-of-speech tags as placeholders in those cases in the first place. German also used to have the elongated *f*, but in the words of the metaphors at least, we can always tell a wrongly parsed lowercase *f* to be an elongated *f* and so we simply have to perform two searches for these: one with *<s>* for 1800 onwards and *<f>* before 1800. Of course we want to search for both over the entire time interval just to make sure we do not miss any occurrences, plus, we would expect that there was a transitional period where some authors and printers still used the elongated *f* while others were already using the regular *s*.

Because books were not published as frequently in the 16th, 17th and 18th century as they are today—sometimes there are gaps of several years without any publications—we have to be critical whether a frequency count based on a postponed publication might actually represent an earlier stage of the language, from when it was originally written. The low publication volume and their infrequency before 1800 makes the frequency counts we get highly unstable and unreliable with huge standard deviations. Since all of our metaphors are already so much less frequent than the most frequent words in English and German—such as determiners, pronouns, and auxiliary verbs, among other function words—we usually do not find them before 1800 at all which creates the false impression in our diachronic corpus analysis, if we were to blindly follow a quantitative approach, that the metaphors originate at the beginning of the 19th century. There are some metaphors that originate in the 20th century (e.g. *love is a battlefield*, *love is a journey*, *The brain is a computer*, *sport(s) is murder*) and we can be quite sure about that because the publication volume in the 20th century is hundreds of times higher than in the 19th century. In the early 19th century, however, there is no remedy against this false impression created by the lower publication volume. Due to these shortcomings of the online Google Ngram Viewer, I had to write my own search engine in Python (van Rossum & Python Software Foundation, 1991) to search the roughly 200 GB worth of raw data.

The main features of this search engine are such that they try to compensate the deficits

in the raw data by workarounds: The algorithm finds all possible lowercase, uppercase, and all-capital letter combinations by converting the raw data always into lowercase and which checks it against the list of search queries that have equally been turned into lowercase while the match is not case-sensitive. In version 2, data files are now sorted alphabetically so that we can be sure that all instances of a word for all years in the corpus are together in one data file, which in turn means we would only need to read that one file. But in version 1 files had an arbitrary size limit so that all the files were about the same size which, however, in turn meant that the tokens per year for an n-gram could be split between separate files. All n-grams were sorted alphabetically, but since the division between files was determined by file size it would often be the case that for a particular n-gram, the earlier centuries would be in one file and at some arbitrary point there would be a cut and the later centuries would be located in another file. So there was no way of knowing in which file a particular alphabetical letter would start and in which file it would end. In effect this meant that for the entire list of metaphors (and their duplicates to anticipate wrong parses) I had to always check against the entire corpus—especially because mis-parses could produce spellings that would put these misspelled tokens 20 or 30 files earlier or later than the tokens of the query that were parsed correctly.

To anticipate these spelling mistakes, i.e. parsing errors, I used an iterative filtering approach, meaning I would first get a rough count by searching not for exact matches of the query but those where the majority of the characters in the string of the n-gram matched the words of the particular metaphor (or its spelling duplicate). This could mean that for a particular metaphor many multiple spellings would have to be checked against all n-gram files as well as paraphrases for metaphors such as *an Englishman's home is his castle, a man's home is his castle, a person's home is his/her castle, my home is my castle*. Whenever such a line contained that rough-matching token, its frequency, the year it appeared in, the number of books and number of pages it appeared on, that line was copied to a separate file specific to that metaphor. This way the algorithm sorted out potential candidates which in the next step would then be checked more closely for potential spelling mistakes. To get as much context as I could from the files I searched among the 5-grams, i.e. strings of 5 words. Since much of the disambiguation cannot be automated I went over the separate files generated in the first step and sorted out cases that were either clearly wrongfully miss-parsed or that were complete matches without spelling variation, leaving only those cases where disambiguation was not possible. The cases that could not be disambiguated were of the *cafe/cafe* variety and were ignored in subsequent filtering. The cases that were complete matches and those that could be disambiguated were then parsed again, this time with stricter matching criteria, and so iteratively the most likely true frequency counts for the metaphors were accumulated. In order to make sure that my search engine worked properly, I used it on the Google Books n-grams US, Fiction, and UK corpora, for which I had used the BYU interface originally. If my search engine worked properly, it should find as many tokens for each metaphor as I had found with the BYU interface. It turned out that because of the iterative filtering in my search engine, not only did it find as many tokens as the BYU interface, but often a few more. I could therefore be sure that my search engine was working properly.

In version 1, the Google n-gram corpora had no labels for text genres and so this renders the question of how to resolve n-gram ambiguities of the *cafe/cafe* variety moot. The Google n-grams Fiction corpus contains fiction books written in English (it contains both fiction books written in British and American English, but Google states that the majority are fiction books

published in the US and so American English) while the Google n-gram UK and US corpus contain all text books written in British and American English respectively, regardless of text genre. For all we know the Fiction n-gram corpus is probably included in the UK and US corpora anyway (Google is not specific on this point) and so a comparison makes no sense. Without a purely non-fiction English corpus there is no baseline against which to compare the Fiction corpus.

I did set a minimum and that minimum is at least 1 token because, as Google states, in the raw data files with the n-grams, n-grams that occurred less than 40 times in the entire corpus were omitted from going through the algorithm that counts their frequency per year because that would have substantially increased the amount of processing time because the number of these highly infrequent items would also increase exponentially and for all we know might not even be real words, but simply nonsense produced by the OCR misspelling words in the scanning process. So even though I set the minimum frequency to 1, effectively that means it is 41.

Since the iterative filtering approach increased processing time tremendously (we are talking about a week for a list of 50 metaphors and their duplicates on dual-core 2 GHz processors) I created frequency counts per year. Summing these raw frequency counts per decade was much faster than if the search engine would have looked up all tokens in all decades for each search query *before* doing the same for the next query.

5.4 Frequency synchronic: The frequency distribution of an entire vocabulary

For the experiments that follow we need to ensure two things when it come to the degree of familiarity of the metaphors: (1) There are thousands of proverbs. We want to pick those that are metaphoric while also having the linguistic form we want, but we also need to make sure that out of all the proverbs that meet these criteria we pick the ones that have the highest chance to be familiar to participants in the experiments. Hence we want English, German, and cross-linguistically shared metaphors that participants will be familiar with. Conversely, when it comes to the novel metaphors created for the experiments, we want to ensure that participants are unlikely to be familiar with them, or better that there is a very high chance that participants have never come across them. (2) We want to ensure that the degree of familiarity of each metaphor type is comparable between English and German: L1 metaphors (English metaphors to English native speakers, German metaphors to German native speakers) should be comparable if they have the same frequency in corpora of the two languages. The same should ideally be the case for shared and novel metaphors. Under the best circumstances, we would also want that all conventional metaphors (the L1 metaphors and the shared ones) are comparable in terms of their familiarity. We therefore need to check that these requirements are met. The reason is that we do not want rating differences in the plausibility judgement task to be the result of some metaphors being more familiar than other metaphors. In other words, if we should find that participants rate metaphors of a certain type or language systematically different then those of another type or language, then we need to be able to check if this difference in rating behaviour could be the result of differences in metaphor familiarity.

We want to make sure that the language-specific metaphors in our experiments, although

they do not appear in the other language, are similarly frequent in use in the two languages. Corpus frequency is an indicator of usage frequency. However, corpus frequencies cannot be compared directly across languages. In order to see that this is so, suppose we wanted to compare corpus frequencies between English and a language whose vocabulary is much smaller than English, whose speakers, however, talk about as much a variety of topics as English speakers do. We immediately know that the speakers of that language would be using the same wordforms in multiple conversational contexts where English speakers would use different wordforms. So simply because the vocabulary of that language is smaller, the frequency of each wordform would, on average, be higher than the frequency of English wordforms. This also means that we could not compare lexical frequency between the two languages. German is known to have a richer inflectional morphology than English. We should therefore expect that one English wordform, which relates to multiple German inflected wordforms, will occur more often than each of the corresponding German wordforms. In that case, since German has more inflected wordforms, it will have more wordforms than English overall in a corpus analysis. Due to the fact that the occurrence of one inflectional form rather than another is constrained by the grammar as well as the discourse context, we cannot necessarily assume that if we tallied the frequency of all German wordforms corresponding to one English wordform that their total frequencies should be comparable. So we would want to express the language-specific frequency of particular wordforms relative to the total number of wordforms of a language and as a function of grammatical constraints.

It is common practice to express corpus frequencies as a ratio: the ratio of the number of tokens of an item of interest over a measure of utterance size, for instance, as the number of tokens per million words of corpus text. Taking the ratio makes sense because it allows us to predict how likely we are to find a particular item, say, how many times, plus or minus a certain error (which we can also predict from corpora), an item will occur on a random page of any random book. We might think that if frequencies are not comparable across corpora and languages that this was because we were using the wrong denominator in the ratio. And we might assume that instead of expressing frequency as a ratio of tokens over utterance length, we should express it as the ratio of tokens over vocabulary size. Intuitively, this might seem like a viable solution to the problem of cross-linguistic variation in vocabulary size, but, as we will see, this approach must *fail* because it falsely assumes that the relationship between frequency and the number of wordforms of a language is linear. We will see that the relationship is non-linear, and the non-linearity also incorporates grammatical constraints. The advantage of the non-linear approach I propose here is that it allows us to find a probability for each lexical item that can be interpreted twofold: as the likelihood of finding the item in a linguistic utterance or discourse of a particular size, which, notice, is in the spirit of the frequency measure already in use; and as the likelihood of any speaker to be familiar with that item. Not only is this probabilistic measure of frequency broader than the measure currently in use, it also addresses the problem of cross-linguistic variation in vocabulary size.

Let us consider another approach to lexical frequency, which illuminates the issues with defining it as a ratio, and, from there, work our way back to the problem of cross-linguistic variation in vocabulary size. Compiling lists, such as the 1,000 most frequent words in English (see Mark Davies' work at <http://corpus.byu.edu/>, for instance), is quite common, where the rank is based on an item's token count per amount of corpus text. Ranking lexical items according to their frequency has lots of practical applications: prioritisation of database access

and information retrieval and the optimisation of encoding in communication technology. Frequent items should be prioritised because they are frequently searched for and their coding must be less complicated than for infrequent items to allow for optimal performance in information retrieval. This rank-based optimisation is also needed to improve the performance of searches in the very corpora we use to find frequencies in the first place. Corpus analysis without this optimisation becomes almost impossible given the enormous size of today's corpora (take, for instance, the ~340 *billion* words worth of text of corpora used in this thesis). However, notice that although rank is assigned according to frequency, we have abstracted away from frequency and compiled a list that, at least to some extent, can be understood probabilistically: the higher an item's rank, the more likely we are to come across that item and the more likely it is that speakers will be familiar with it. Statistical significance is commonly assessed probabilistically. The probabilistic nature of the ranking naturally lends itself to judging the significance of frequency differences. But we have also lost something through ranking: the distance of a given rank to its neighbouring ranks.

Another feature of ranking lexical items is that it is quite possible for multiple items to share the same rank. So we need not only ask "How great is the distance between neighbouring ranks?" but also "How many wordforms share the same rank?" The distance and number of wordforms per rank is effectively a measure of density: the more wordforms there are per rank, the higher the density; the smaller the distance to neighbouring ranks, the higher the density. It is also quite intuitively obvious that, *a priori*, we have no reason to assume that the density should be constant as frequency increases. As we will see, the density is, indeed, not constant and so judging the significance of frequency differences of any two lexical items, ultimately depends on the rate of change in this density. From these contemplations on frequency-wordform density we can see that instead of asking "How frequent is a particular lexical item," we should be asking "How many wordforms have the same frequency density." Since the density depends on two factors, frequency closeness and the number of wordforms per frequency rank, we can split the question into two parts: first, we will find the distribution to the question "How many wordforms are there per frequency rank" and then we will permute this distribution by asking "How close is each frequency rank to its neighbours," thereby giving us the rate of change in density as a function of frequency. We want the density distribution to be a function of frequency, because, ultimately, we want to be able to evaluate the *frequency* of particular lexical items against this frequency-wordform density distribution of the entire vocabulary.

In order to answer the question "How many wordforms are there per frequency rank," I designed a Python programme to search the Google n-gram corpora for British English, American English, and German. The n-grams in these corpora are effectively wordforms. Larger corpora give better estimates of wordforms' frequency (token count per million words worth of corpus text). The Google n-gram corpora seemed like the logical choice, since they contain more material than the part-of-speech tagged corpora combined (British National Corpus, Corpus of Contemporary American English, Corpus of Historical American English, and the German COSMAS-II corpus project). We want to use the most recent data for this. Google discontinued scanning and adding more texts to the corpora in 2009 for English and 2008 for German, meaning that there was less text material than in the years before that. Hence, I used the year before that for the respective languages, 2008 for English and 2007 for German. Non-automated iterative filtering was simply out of the question for this task and automating it would also not have been advisable given the context-dependent nature of the disambiguation process. Even

if we only looked at the wordforms of just one year, we would have to manually check around 3.6 million wordforms for British and American English and German. However, there are also drawbacks to using n-gram rather than part-of-speech tagged corpora to estimate the number of wordforms. Then again we are interested in the familiarity of metaphors' linguistic form, so the restriction to wordforms does also work in our favour.

The limitations, generally speaking, fall into two categories: We either overestimate the number of lexemes (the size of the vocabulary) by counting wordforms or we underestimate them. The most problematic cases are those where the same wordform corresponds to different lexemes, e.g. nouns and verbs, or tensed wordforms that correspond to different verbs such as *lie* as in *telling a lie* and *to lie down*. These problem cases make the number of wordforms seem smaller than it really is because they conflate different grammatical words and/or lexemes. Spelling errors of the *cafes/cafes* variety would also be treated as different wordforms. It also means that whenever two wordforms are homographs we treat them as one and the same wordform, e.g. *bank* as (1) an institution where money is kept or (2) a river bank, the side of a river, or where different case inflections in German are isomorphic and homographs, e.g. *der* (*the*) can be nominative singular or genitive plural for masculine nouns; genitive singular, dative singular, or genitive plural for feminine nouns; or it can be genitive plural of neuter nouns. The frequency of function words depends on the frequency of content words: for instance, how many tokens of determiners we find in the corpus depends on the number of nouns in it. Similarly for auxiliary verbs: since auxiliary verbs require a main/lexical verb, their frequency is a multiple of the number of main verbs in the corpus.

Many verbs and nouns in English have the same morphology and spelling, i.e. they are isomorphic and homographs, such as *iron*, *drive*, and *move*, which can be both nouns and verbs. In German corpora, there will be many homographs that really ought to correspond to different lexemes. An example for the case syncretism within the definite determiners is given above. Verbs will have a lot more inflectional forms, many of which will also be subject to formal syncretism without part-of-speech tags. Some nouns, such as *Hahn* correspond to two different lexemes (*Hahn* can refer to a rooster/male chicken or a valve in plumbing), but have completely identical inflectional paradigms. There are also cases in which one wordform relates to two different senses: for instance, *random*, which can either describe (a) spontaneous, impulsive, or unexpected behaviour or (b) that something is selected at random, or *utopia*, which can either refer to (a) a dream or goal that cannot be realised, but is highly admirable or (b) a foolish dream or goal that should not be pursued.

In cases where multiple wordforms correspond to the same lexeme, we overestimate the size of the vocabulary by counting wordforms. All regular verbs in English, such as *walk*, have four wordforms (*walk*, *walks*, *walked*, *walking*) that correspond to only one lexeme and all irregular verbs, such as *go*, have five wordforms (*go*, *goes*, *went*, *gone*, *going*); and these wordforms correspond to at least six different grammatical forms (more if we consider imperative and subjunctive mood to be separate forms from the plain form, as in *They suggested his comment be put on record* or *Be careful*). Hence for all regular verbs we would always treat past tense and participle forms ending in *-ed* as one and the same. Similarly, homographic and homophonous nouns. *People*, for instance, corresponds to two distinct lexemes: (1) a group of individuals versus (2) the citizenry of a country or any other territory that consider themselves a unit. Also the plural form of the former is the same as the singular of the latter. Alternative spellings and/or inflectional forms, e.g. *formula* (sg) and its two alternative plural forms *formulae* (pl)

and *formulas* (pl) or the two plurals for *fish* (sg), *fish* (pl) and *fishes* (pl) as in *sleeping with the fishes* or when referring to the astrological sign (in which case one might argue they would be two different lexemes or at least senses again, as with *people*).

We might be able to devise an algorithm that can check for parsing errors in regular inflectional forms, where these errors follow predictable patterns, but if we consider the same approach for derivational morphology, we would quickly run into non-trivial decisions which we either do not want to leave to an automated parsing algorithm (which might get it wrong) or which cannot be decided because they are both morphologically and semantically ambiguous: for instance, *un-lock-able* is both morphologically and semantically ambiguous; it can mean either (a) something that can be unlocked or (b) something that can never be locked or cannot be locked at all. In certain cases, parsing errors would be indistinguishable from morphological derivatives. For instance, *indispensable* could be the morphological derivative *in-dispensable*, but it could also be a parsing error, the preposition *in* and the adjective *dispensable* in a prepositional phrase such as *in dispensable cups*; and due to the lack of part-of-speech tags disambiguation is impossible in these cases. There are many exceptions to the way that derivational affixation is compositionally predictable. *Alive* and *dead*, for instance, are antonyms, but *undead* is not simply the negated/inverted meaning of *dead* as it is not synonymous with *alive*. It means *immortal*—or at least one of the senses of *immortal* as it can either mean (1) cannot die and will live for ever or (2) has died, but is still animate. The limitations imposed by having to use wordforms as a basis for the estimate means that we need to take it with a grain of salt, but it also means that it is not entirely impossible to establish such an estimate.

Let us return to assigning rank by frequency. That grammar would dictate the number of wordforms per frequency rank seems like a reasonable assumption. Consider, for instance, that articles such as *a/an*, *the*, or *this* occur in the presence of nouns such as *dissertation* or *paper* (depending on our definition of “article,” of course; cf. the DP hypothesis). So the frequency with which articles occur depends on the frequency with which nouns are used. The frequency of auxiliary verbs, such as *be* and *have*, and modals, such as *would*, *could*, and *should*, depends on the frequency of main verbs such as *read*, *write*, or *submit* because the grammar constrains their occurrence. In a way similar to the co-occurrence of articles & nouns and auxiliary & main verbs, the frequency of prepositions such as *in*, *on*, or *at* depends on the wordforms that co-occur with them in prepositional phrases. We find a general pattern: articles are more frequent than nouns, auxiliary verbs are more frequent than main verbs, and prepositions more frequent than the other wordforms in prepositional phrases, which are often nouns. So function words such as articles, auxiliary verbs, and prepositions are generally more frequent than content words such as nouns and main verbs. This affects the distance between neighbouring ranks as well because the occurrence of content words determines how many times function words can occur. This naturally leads to the distance between neighbouring ranks increasing exponentially. Notice that the number of wordforms per rank and distance between ranks reflects semantic complexity as well. Content words are rich in meaning and polysemy, with highly contextual Fregean senses, whereas function words, compared to content words, are often much more restricted in meaning and if they display contextual polysemy, their inferential interpretation must make reference to their phrasal constituents, which are, of course, content words. So the frequency of function words multiplies by a factor and this factor is the frequency of their respective phrasal constituents. Given these grammatical constraints, we might not expect the number of wordforms per frequency rank to be easily predictable; however, as we will see, the

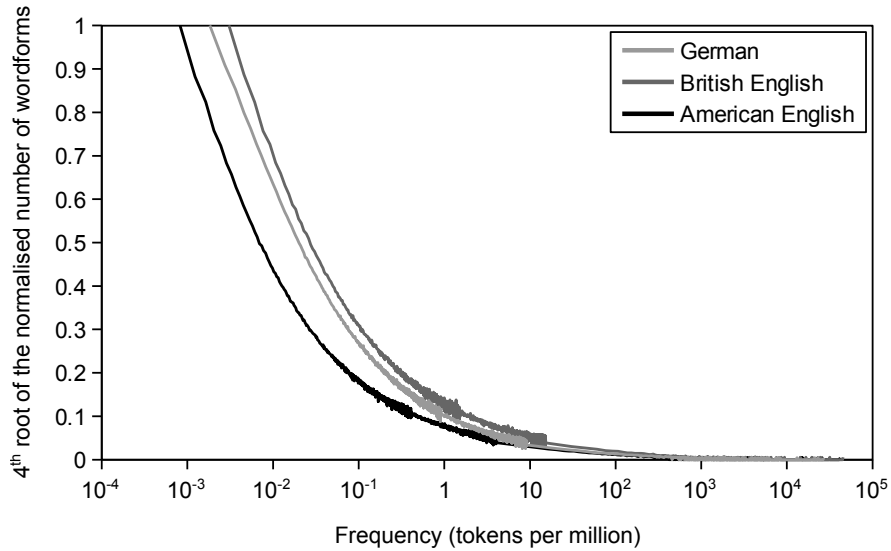


Figure 5.1: The raw distribution of the number of wordforms per frequency $n(f)$ for the vocabularies of British English, American English, and German (21,908 data points). For visualisation purposes the y-axis is normalised and scaled to the 4th root.

number of wordforms per rank decreases exponentially as a function of frequency.

We get what looks like a power function, shown in Figure (5.1). In a regular power function, there is only one exponential function. However, in our case we know that frequency increases exponentially and the number of wordforms per frequency decreases exponentially with frequency. So in our case, what looks like “one” power function is actually two exponential functions intertwined, one monotonically increasing, the other decreasing. We can therefore see that the “familiarity probability of a particular lexical item” is not simply its frequency, but its frequency *relative* to the frequency of the rest of the vocabulary. In order to determine an item’s familiarity probability in this relative sense, we need to disentangle the two exponential functions, $f = x^b$ and $n(f) = y^c$. The problem, however, is that we do not know the exponent of either one of the two functions, b and c . We do know that $\log_x(x^b) = b$ and $\log_y(y^c) = c$. Although we know neither b nor c , we know that for all logarithms of the same monotonicity, changing the base of the logarithm from b to c , the two exponents we are looking for, is

$$\log_c Z = (\log_c b) \log_b Z. \quad (5.1)$$

The number of degrees of freedom in finding b and c thus reduces by one degree, which means that if we can find either b or c , we can use Equation (5.1) to find the other. Later, we will use the conditions of the cumulative probability function to do just that. From Equation (5.1) we can see that in order to separate the two exponential functions, we would have to divide them to find their conversion factor, $\log_c b$. All monotonically *increasing* exponential functions are proportional to one another and all monotonically *decreasing* exponential functions are proportional to one another. This is also why the choice of base of the logarithmic scales in the graphs is arbitrary. The exponential function, $f = x^b$, that belongs to the frequency scale is monotonically increasing. The exponential function, $n(f) = y^c$, that corresponds to the

number of wordforms n per frequency f is monotonically decreasing with respect to frequency. Therefore, the two functions are inversely proportional, $n(f) \propto 1/f$: as frequency f increases, the number of wordforms n decreases. With $\log_c b$ as proportionality factor, C , the proportionality becomes an equality, which in turn becomes

$$C = \frac{n(f)}{1/f} \quad (5.2)$$

when we solve for C , and which we can simply write as $n(f) \times f$. Equation (5.2) also shows us that it is irrelevant which way we divide the two functions; for convenience, however, it is useful to plot the functions relative to frequency. Since the two exponential functions are related to each other through frequency, we can think of C as a frequency density coefficient. If we count the number of wordforms per frequency in this way we get the graphs shown in Figure (5.1) for the vocabularies of British English, American English, and German. Shown is the number of wordforms n for a particular frequency band f , normalised, i.e. divided by the largest number of wordforms of any frequency band n_{max} , given in Equation (5.3). For visualisation purposes, this function is scaled to the fourth root, $\sqrt[4]{n(f)/n_{max}}$, in Figure (5.1).

$$f_{raw}(x) = n(f) \quad (5.3)$$

We need to convert the raw power function (Equation 5.3) into a frequency-wordform density distribution (Equation 5.4), i.e. we need to transform the number of wordforms per frequency function $n(f)$ into a function of the density per frequency, $n(f) \times f$. For the conversion we use the function $n(f)$, not the function $\sqrt[4]{n(f)/n_{max}}$ which is given in Figure (5.1) for visualisation purposes only. Multiplying $n(f)$ with the frequency f has turned the y-axis into a logarithmic scale. In order to transform it back to a linear scale we use $n(f) \times f$ as the exponent of a factor, b . Factor b has to be chosen such that in the cumulative distribution function in Equations (5.5) and (5.6) where the mean already is at 0.5 on the x-axis, it also passes through 0.5 on the y-axis (see Figure 5.3) because this is a requirement for all cumulative probability distributions and b has to be adjusted for each language individually. Lastly, we divide by b to normalise the function once more. This gives us the density function f_X for a particular frequency band x shown in Figure (5.2).

$$f_X(x) = \frac{b^{(n_x(f_x) \times f_x)}}{b} \quad (5.4)$$

The resulting density function, as we can see from Figure (5.2), is a binomial function with the characteristic bell shape of a Gaussian normal distribution. In any binomial probability mass function $f_X(x) = \Pr(X = x)$ of a random variable the probability for a particular x -value must be between 0 and 1 and the sum of all probabilities is exactly 1, $\sum_{i=1}^n f(x_i) = 1$. By normalising Equation (5.4) we effectively interpret the density of wordforms $n(f) \times f$ per frequency band x as their probability $\Pr(X \leq x)$. For any such binomial distribution $f_X(x) = \Pr(X = x)$ there is a cumulative probability distribution $F_X(x) = \Pr(X \leq x)$ associated with it (shown in Figure 5.3).

The cumulative distribution provides us with the probabilities per frequency that we need in order to assess the likely familiarity of language-specific and cross-linguistically shared metaphors. These are the probabilities given in brackets in Table (5.2). The cumulative probability distribution is derived from the density function by interpreting densities as probabilities:

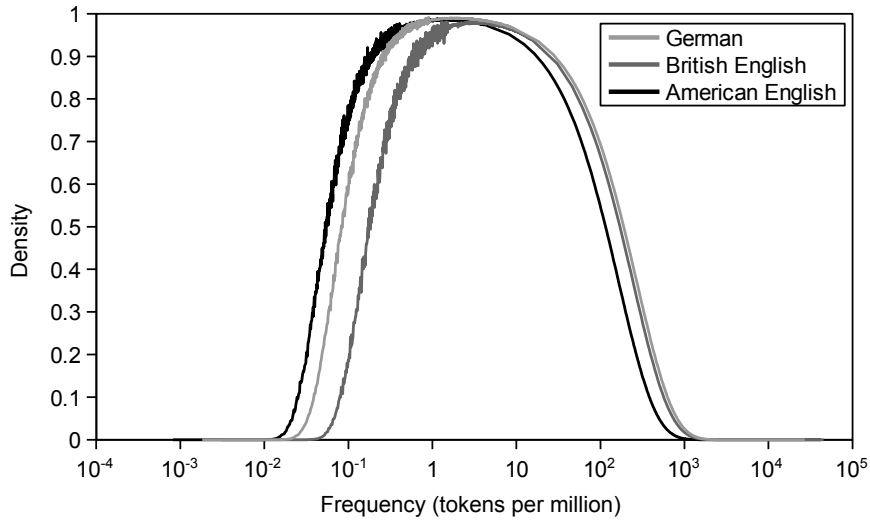


Figure 5.2: The probability mass function $f_X(x) = \Pr(X = x)$ for the vocabularies of British English, American English, and German.

higher density, for a particular frequency band, means higher probability, lower density means lower probability. For any particular frequency band x , its cumulative probability $\Pr(X \leq x)$ is its own probability $p(x)$ plus the sum of all probabilities of frequency bands x_i with a lower frequency (x_1, x_2, x_3, \dots) $< x$:

$$F_X(x) = \sum_{x_i \leq x} \Pr(X \leq x_i) = \sum_{x_i \leq x} p(x_i) \quad (5.5)$$

For a particular frequency band x we therefore sum its density $f_X(x)$ that we get from Equation (5.4) with the density of all lower frequency bands $x_i \leq x$ and divide by the sum of all densities $\sum_{j=1}^n f_X(x_j)$ (where n is the number of all densities, i.e. frequency bands) which gives us the following equation for the cumulative probability distribution:

$$F_X(x) = \sum_{x_i \leq x} p(x_i) = \sum_{x_i \leq x} \left(\frac{f_X(x_i)}{\sum_{j=1}^n f_X(x_j)} \right) \quad (5.6)$$

The resulting cumulative probability distribution is shown in Figure (5.3). The x-axis is normalised, i.e. the frequency of the most frequent wordform in English and German, the definite determiner *the*, *der*, *die*, *das*, is equal to 1. This is in order to demonstrate that the mean of the languages' distributions falls on $x = 0.5$. Notice the steep inclination of the slopes and how their inclination is quite similar among the languages. There is no intrinsic reason in the derivation of the cumulative probability distribution that this should be the case. Therefore their cross-linguistic similarity comes as a surprise. Why the slopes have this inclination is beyond the scope of this thesis, but I have included two Gaussian normal distributions in order to show an important conclusion of deriving the cumulative probability distribution as a function of corpus frequencies: if one tries to fit a normal distribution to the upper half of the cumulative distributions of British English, American English, and German, that same normal

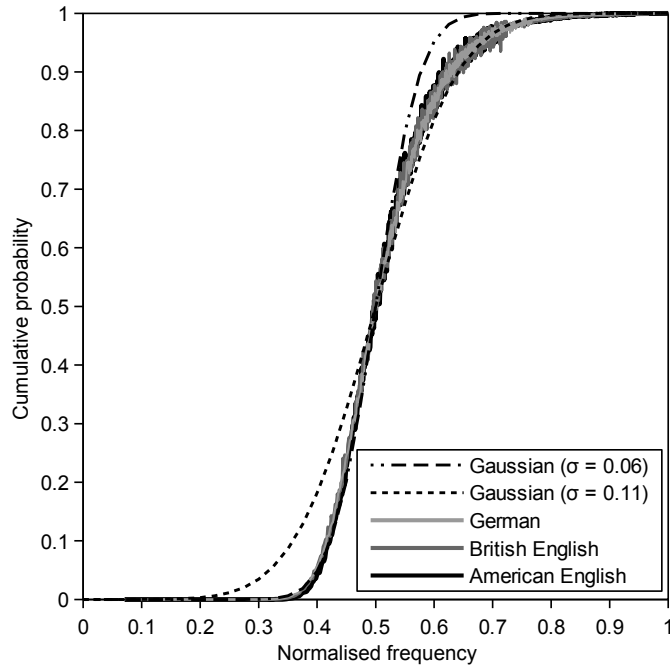


Figure 5.3: The cumulative probability distribution function $F_X(x) = \Pr(X \leq x)$ for the vocabularies of British English, American English, and German.

distribution does not fit well with the lower half of the cumulative distributions. Vice versa, if one tries to fit a normal distribution to the lower half of the cumulative distributions, that normal distribution will not fit the upper half of the cumulative distributions. This shows that it is not appropriate to use a fitted Gaussian normal distribution in place of the languages' actual frequency density and cumulative probability distributions. If we did so then, judging from the offset visible in Figure (5.3) between the two Gaussian curves and the languages' cumulative probability distributions, we would have derived probabilities for the likelihood of language-specific and cross-linguistically shared metaphors to be familiar that were up to $\Delta p \approx .1$ off from the true cumulative probabilities. This offset is quite substantial given the fact that the frequency differences between the language-specific metaphors and cross-linguistically shared metaphors that we are interested in evaluating statistically gives us the following differences in cumulative probability: $\Delta p = .0315$ for American English, $.3065$ for British English, and $.1585$ for German. The offset of $\Delta p \approx .1$ is quite big compared to these differences in cumulative probability and would have diminished the reliability of the statistical analysis. It is therefore important to stress that it is *not appropriate* to use fitted normal distributions instead of the actual frequency density and cumulative probability distributions.

5.4.1 Cross-linguistic comparability of metaphors' corpus frequency and familiarity probabilities

With the help of British English, American English, and German corpora we got the following metaphor frequencies per metaphor type and language variety, shown in Table (5.2). For a complete list of the exact token counts in the individual corpora, see Table (A.3) in Appendix A.

Table 5.2: Corpus frequencies F (and cumulative probabilities p) per metaphor type averaged across the nine English and German corpora. Columns are American English (AE), British English (BE), and German (G); rows are the metaphor types: English-specific (E), German-specific (G), cross-linguistically shared (S), and novel (N).

$F(p)$	AE	BE	G
E	.24 (.46)	.11 (.12)	0
G	0	0	.11 (.32)
S	.55 (.49)	.43 (.43)	.46 (.48)
N	0	0	0

In Table (5.2), the frequency is the number of tokens per million words in the corpus and the p -value is the cumulative probability, i.e. the probability to come across a metaphor compared to the entire vocabulary similar to a z -score (the vocabularies of AE, BE, and German were assessed independently). Take the probability of shared metaphors in the American English (AE) corpora as an example: A cumulative probability of .98 can be interpreted to mean that participants will be 49% likely to be familiar with the metaphor because the chance to have come across a shared metaphor out of the sum of the chances to come across other vocabulary items is 49%. In the next section I will explain in detail how I calculated these cumulative probabilities. The frequencies per metaphors type in American English (AE) are the average metaphor frequencies of each type in the COCA, COHA, Google Books n-grams US, and n-grams Fiction (according to Google’s own statement, English books in the category fiction are mostly those published in the US). For British English (BE), the frequencies shown in the table are the average frequencies in the BNC and Google Books n-grams UK. The metaphor frequencies for German (G) are the averages over the three German corpora: COSMAS2-W, COSMAS2-H, and Google Books n-grams German.

Due to the fact that frequencies increase exponentially, the significance of differences and its statistical evaluation needs to grow along with it. The more frequent the groups of items that are compared are, the more pronounced the difference in frequency has to be in order to amount to the same statistical significance. In the previous section I proposed a method to estimate the vocabulary-wide frequency variation by converting the raw number of wordforms per frequency into a density function, i.e. a function of the number of wordforms per frequency and closeness to the next frequency bands of lower and higher frequency. This density function has the properties of a probability mass function $f_X(x) = \Pr(X = x)$ of a discrete random variable. For any probability mass function there is a cumulative probability distribution $F_X(x) = \Pr(X \leq x)$ associated with it. I suggested that we can interpret the cumulative probabilities that we derive from the cumulative probability distribution as the likelihood of a lexical item, such as a metaphor, to be familiar.

When we take the frequency distribution of the entire vocabulary into account in this way, we can gauge the significance of the frequency differences between L1 and shared metaphors by subtracting the cumulative probability of the L1 metaphors from the cumulative probability of the shared metaphors (see the p -value in brackets in Table 5.2). We get the following differences: $\Delta p = .032$ for American English, .307 for British English, and .159 for German. For American

English $f(S)$ is likely similar to $f(L1)$. For German it is likely that the difference $\Delta p = .159$ means they are comparable. Δp is much higher for British English than American English or German, which in part is probably owed to the fact that (1) the size of the British English corpora was much smaller than those for American English and German and so the estimation of the frequency distribution of the entire British English vocabulary is less reliable and (2) metaphors are systematically less frequent in the British English corpora than they are in the American English and German corpora (see Table 5.2)—most likely because of the ratio of text genres in them: the British English corpora have more newspaper articles in them, whereas the American English and German corpora have more literary and poetic texts. In order to adequately gauge the significance of the cross-linguistic comparability of frequency, we first need a way to determine the amount of error in measuring frequency. In Section (5.6), I suggest a method to determine the corpus-specific measurement inaccuracy. In Section (5.7), I will then discuss how to gauge the substantive and statistical significance of frequency differences.

5.4.2 Lexical frequency of metaphors' source and target concepts

We can use the same method of deriving cumulative probabilities from corpus frequency for the words that make up the metaphor sources and targets. We find that the words are more frequent than the average frequency of cross-linguistically shared metaphors. Hence, if experimental participants are likely to be familiar with the shared metaphors, they are most definitely likely to be familiar with the words in them. This analysis also confirms that the choice of words does not include rare poetic terms (maybe apart from *froth*). For a complete list, see Table (A.2) in Appendix A.

Of the words that the metaphors are composed of, the English words *dagger*, *sponge*, *volcano*, *squirrel*, *battlefield*, and *froth* (in descending frequency) and the German words *Schlachtfeld* (*battlefield*), *Dolch* (*dagger*), *Vulkan* (*volcano*), *Schaum* (*froth*), *Schwamm* (*sponge*), and *Eichhörnchen* (*squirrel*) are as frequent. All other words in the metaphors have a higher frequency. Thus all of the words are highly frequent and highly familiar words and it is unlikely that they should be the cause of differences in ratings of the metaphors' meaningfulness. Notice in Table (A.2) that we get the exact same words at the lower end of the word list arranged by frequency in English and German. If metaphors' source and target concepts are frequent, their Fregean senses are frequent, too. Variation in the frequency of their senses is variation about the average frequency of their host concepts. If metaphors' source and target concepts are more frequent than metaphors' idiomatic meanings, then the source and target concepts' senses are more frequent than their idiomatic meanings as well.

We can also use this knowledge about the words in the metaphors to put the cumulative probabilities of the metaphors themselves into perspective. Recall from Table (5.2) that, on average, cross-linguistically shared metaphors have a cumulative probability of .47 and language-specific metaphors have a cumulative probability of .30. Earlier I said that we can think of the cumulative probability as the likelihood of the metaphors to be familiar, but *how familiar* is .47? We can also think of cumulative probability as the chance to come across a metaphor. If it is close to .5 then every other sentence could be a shared metaphor, but this will only be true if we look at a very large number of sentences. Now that we have assigned cumulative probabilities to the words in the metaphors we can also call on our linguistic intuitions about familiarity: cross-linguistically shared metaphors, such as *time is money* and *silence is*

golden, should be as familiar to us as the words *pay day*, *froth*, *election day*, *squirrel*, and *dagger* because they have similar cumulative probabilities, and language-specific metaphors such as *talk is cheap* and *dreams are froth* should be slightly less familiar to us. The language-specific metaphors have three quite different cumulative probabilities in each language and language variety: they have the lowest probability in British English, but we can say that they should be as familiar to us as words such as *consenting*, *snowballing*, *non-refundable*, or *to rebuff* (further examples taken from the BNC). Language-specific metaphors has a cumulative probability of .32 in German corpora and should thus be as familiar to us as *snowman*, *to cycle* or *boatload* in English (further examples taken from the BNC and COCA). Language-specific metaphors have the highest probability according to the American English corpora with a probability of .46 and should thus be as familiar to us as words such as *ejecting*, *disassembly*, *squeaks*, or *bombardments* (further examples taken from the COCA). Hence we can draw on our familiarity intuitions to make sense of the cumulative probabilities of the cross-linguistically shared and language-specific metaphors.

5.5 Frequency diachronic: stability over time

Once we have found metaphors that are frequent enough in corpora so that we can assume that speakers should be familiar with them, we need to ensure that these frequencies are stable over time or—if anything—show an increasing trend and we would like this stability and/or trends to be the case for at least the last 30 to 40 years within both English and German so that we can assume that throughout the lifetime of our average participant, he or she has had a stable chance to encounter the metaphors and familiarise themselves with them.

Take the diachronic development of the cross-linguistically shared metaphor *silence is golden* in English and German shown in Figure (5.4) as an example of diachronic frequency changes that we typically find for many of the metaphoric proverbs under investigation here. Our main focus is on the last 30 to 40 years because that is the age of the average participant in the experiments, but we can trace many of the language-specific and cross-linguistically shared metaphor back much further than that. Interestingly, for *silence is golden* find that during the 19th century, the phrasing *silence is gold* co-existed with the phrasing *silence is golden* in English. In German, we only find the phrasing with the noun *gold* and never with an adjective, neither *golden* (G. *golden*) nor *gilded* (G. *vergoldet*) nor *golden* in diminutive form (G. *goldig*). Figure 5.4 shows the diachronic co-development of the two alternative phrasings in English that ultimately resulted in *silence is golden* becoming the predominant, conventional form used today. I think the historic interrelatedness of the two phrasings gives us good reason to believe that the metaphor has the same idiomatic meaning in both English and German. This co-development can roughly be divided into three consecutive phases.

In the first phase, the two alternative phrasings with the noun *gold* and the adjective *golden* co-exist and begin to increase in usage frequency between 1840 and 1870. In this time period, the noun-variant, *silence is gold*, sees about the same frequency increase in English and German, while the adjective variant, *silence is golden*, gains more traction in English than the noun variant. In German, the adjective variant is never established: None of the possible synonyms of G. *golden*, *vergoldet*, or *goldig* are found in any of the three German corpora.

In English, the adjective variant continues its rising trend between 1870 and 1890 and at the same time the noun variant already begins to decline in frequency. This is the second

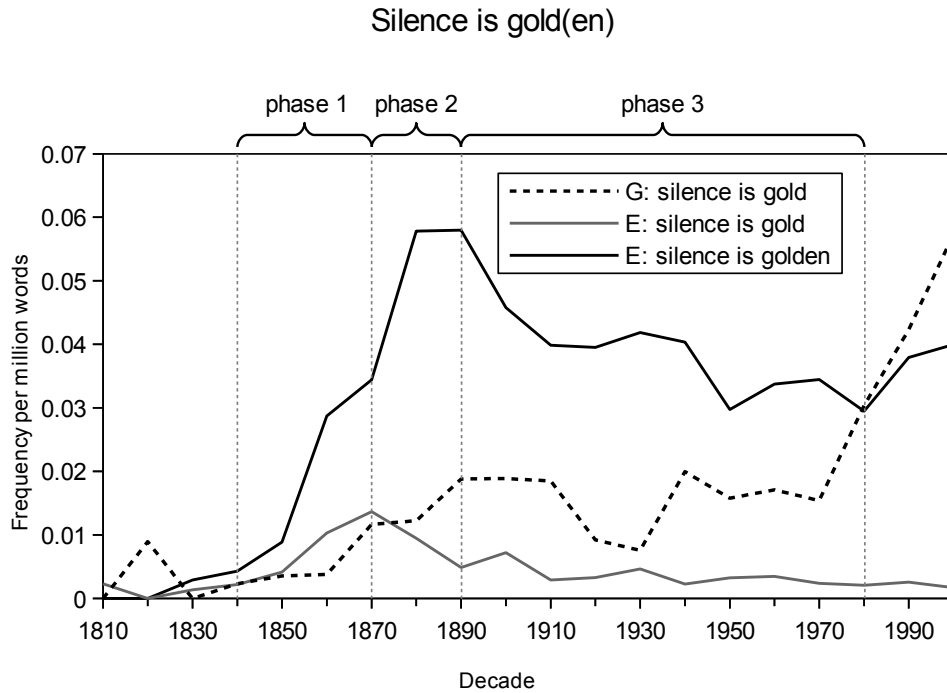


Figure 5.4: Diachronic frequency trajectories between the year 1810 and 2000 for *silence is golden* (Adj) and *silence is gold* (N) in English compared to the trajectory of *silence is gold* (N) in German.

phase. There are two possible explanations here: (1) English speakers see no need to retain both alternative phrasings and in a conscious or unconscious attempt to regularise language the evolutionary bottleneck in trans-generational transmission favours the variant that is already more frequent. In German, there is no competing alternative to the noun variant and so—despite its lower frequency compared to the now preferred English phrasing of the metaphor—the noun variant becomes the established, conventional form of the metaphor in German. (2) The increasing contrast in frequency between the two variants in English has nothing to do with the effects of evolutionary cultural selection. Speakers of English between 1870 and 1890 are familiar with both forms, but they disprefer one—either as a shift in usage within the whole speech community (language change) or as the cumulative result of language change in speakers’ personal linguistic development throughout their lifetime. At the time, we find the noun variant has mostly disappeared from general discourse and almost exclusively occurs in literature, i.e. the Google Books English Fiction corpus (Google, 2012). It could simply be the case that speakers at the time deemed the noun variant more and more a literary or poetic expression, a creative paraphrase to the “common” way of saying *silence is golden*.

After the adjective variant has established itself as the dominant, conventional phrasing of the metaphor, its usage frequency declines over the next 100 years, the third phase. The frequency of the noun variant continues to decline—ever so slightly—but does not disappear completely from the English corpora. However, it is mostly found in corpora of American

English and we might want to hypothesise that this could be due an influx of German immigrants to the United States at the time. In German, on the other hand, notice how the usage frequency of the metaphor—more or less—continues to climb, uncontested by any alternative variants and notice how it seems that the frequencies of both variants decline in English after one variant has established itself as the dominant, conventional form. This is one example of diachronic change in a metaphor’s frequency. Further examples can be found in Figures A.1, A.2, and A.3 in Appendix A.

We would assume that participants are not only aware of how common in usage certain metaphoric proverbs are, but also whether their usage is changing diachronically. We would therefore be interested not to select metaphors for the experiments that, during participants’ average lifetime, have decreased in frequency, but only metaphors that have either remained stable in frequency or increased in frequency. Over the last 30 years, the mean age of participants, none of the experimental metaphors have decreased in frequency. Two cross-linguistically shared metaphors, *knowledge is power* and *time is money*, showed an increasing trend in both English and German corpora and the English-specific metaphor *a home is a castle* (including its various alternative wordings such as *an Englishman’s home is his castle*, *a man’s home is his castle*, *a person’s home is his/her castle*, *my home is my castle*) showed an increasing trend. All other metaphors have stable frequencies (see Table A.4).

5.6 How do we know zero frequency means zero?

In this part, I ask how can we be certain that if, for instance, we have classified a particular metaphor as, say, English-specific because it only occurs in English corpora and never in German corpora, that since the chance to encounter infrequent metaphors depends on the size of the corpus, *how can we be sure that we would not find any tokens of the metaphor in German corpora* (thereby changing its classification from English-specific to cross-linguistically shared) *if only the German corpora were larger in size?* We can never be absolutely *certain*, but we can quantify how much the spread of the metaphor frequencies we have gathered from the corpora *reduces our uncertainty* about this concern. What we want to do to answer the question is this: We know the language-specific metaphors are less frequent than the cross-linguistically shared metaphors, so they are in more danger of being re-classified. At the same time we know that the frequencies of particular language-specific metaphors vary about the mean values given in Table (5.2). We also know that the reliability of these individual frequencies (the accuracy of the measurement) also varies because of the way that metaphors naturally occur across texts and text genres and of how we sample from this real distribution by selecting texts to be included in corpora. In other words, the variation in measurement (inverse of its accuracy) is added to the linguistic variation so that the variation that we actually get in results is the compound variation of the two. Hence, in order to quantify the reduction of our uncertainty we need to measure the difference between zero and (the variation of the language-specific metaphors + the variation in measurement itself). Basically, if the variation (language-specific metaphor + measurement accuracy) is so large that we sometimes expect negative token counts, we are in trouble. However, that is not the case for my corpus analysis, so the metaphor classification is safe. We also need them so that when we analyse our experimental results later on and find differences in metaphor ratings, we have a way of determining whether those differences could have been caused by having less familiar (less frequent) words in metaphors with lower

plausibility ratings.

That the variation in measurement and the variation of the variable we are trying to measure compound in this way is a well-known fact. The real difficulty is to establish a good estimate of the variation of measurement (measurement accuracy) for the corpora. For this we need to measure the fluctuation of items in the corpus that meet two criteria: (1) although their frequency varies, the mean of this variation remains the same (the more stable the mean, the better the estimate of a corpus's measurement accuracy), and (2) they have a very high frequency overall because this means that we can measure their frequency fluctuation more accurately (the more frequent they are overall, the more resolved our measure of their variation is). Essentially, if we had an item in a corpus whose mean frequency remained perfectly constant at all times, then the breadth of the variation about this mean should be the variation in measurement itself. However, all linguistic items will not come anywhere near this ideal because their frequencies vary under the influence of some linguistic variation or other. A classical example of linguistic variation and one that speaks directly to the reliability of our diachronic corpus analysis is that of diachronic change. Earlier in this chapter I discussed the diachronic frequency change of the metaphor *silence is golden* and there are many more examples in Appendix A (Figures A.1, A.2, A.3). Another type of linguistic variation that directly affects our analysis is the variation in the distribution of words that any piece of writing exhibits by virtue of the grammar itself: when and where, say, a form of *be* occurs is not random in the same sense as, say, a specific chemical compound or mineral occurs in a piece of rock. Similarly, when and where a particular noun, say *metaphor*, occurs in a text is partly governed by the grammar and partly the author's choice. The topic of the text, its genre, and stylistic choices such as alternating between synonyms of a term in order to avoid repetition, are all extensions of rational choices made by the author that heavily influence the frequency with which we will find particular words in the corpus.

But what if we could estimate the variation of measurement based on an item in the corpus that is not a word? Because the Google corpora are not part-of-speech tagged, the n-grams also happen to include punctuation marks. I propose that we use punctuation marks to estimate the variation of measurement. Of course, they also follow certain orthographic-grammatical rules, e.g. full stops mark the end of sentences, but may also appear in abbreviations such as Fig. 1 or in the initials of names, but they are (1) highly frequent, much more frequent than most words, and (2) the number of contexts in which they can appear are much more constrained and thus far smaller than for any word. This makes them come as close to the ideal benchmark item as we can probably hope. The Google n-gram corpora are the largest ones I used for the synchronic and diachronic frequency analysis. So if we are worried whether those novel metaphors that now show zero frequency could appear in corpora that were larger, if our estimate of the measurement accuracy suggests that this is unlikely to be the case for the largest corpora in our analysis, then it will be even more unlikely for the smaller ones.

In Figure (5.5), I have plotted the number of words per full stop over the entire diachronic period available from the corpora, 1564 to 2008/2009 (recall, the German corpora end in 2008). Note that so far I have always given frequencies as the number of tokens per million words. The definite determiner in English and German occurred between 27,039 and 45,539 times per million words. In Figure (5.5), we see that most often there are 20 words per full stop and the variation over time is such that it sometimes tends to be greater but it hardly ever falls below 20 words. So given the size of the English and German Google corpora we can therefore put

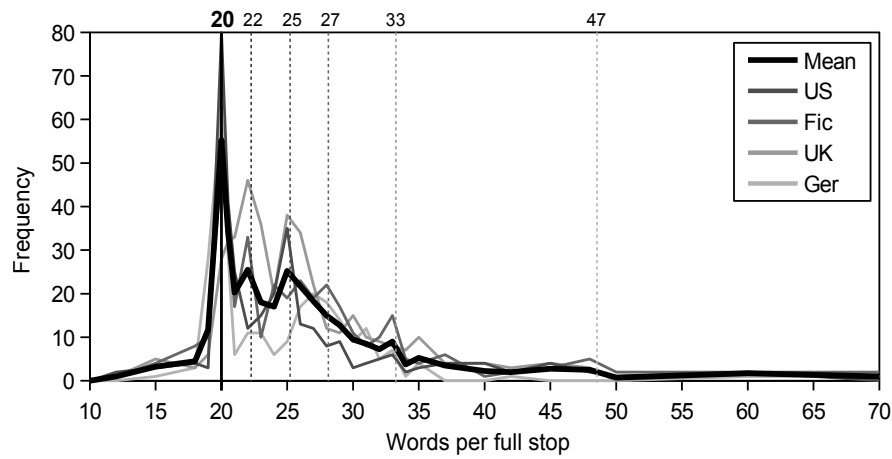


Figure 5.5: Histogram for the number of words per full stop between the year 1564 and 2008 for the Google Books n-gram corpora US, English fiction (mostly AE), UK, and German.

the number of full stops into perspective and say that their frequency is around 49,169 times per million, so about as frequent as the definite determiner. We can therefore be sure that we have met the high-frequency requirement. From Figure (5.6), we can see that we also meet the small fluctuation of the mean over time: the black line shows the trimmed mean over time, between the year 1564 and 2008. The mean is trimmed, i.e. for each decade it shows where 95% of the frequency population lie. Above and below I have plotted the frequency variation about the mean in grey. We see that (1) the variation about the mean is much greater before 1800. This is because of the low volume of texts available for this time frame. We see (2) that in the middle of the 19th century and towards the end of the century, the number of words per full stop drops from around 25 before mid-century to around 20, which is the value that most prominently stood out in Figure (5.5). And (3) we see that for the last 200 years the variation about the mean is very low, in fact, around 1 standard deviation or less. This means that the standard error, the standard deviation divided by the square root of the number of samples (frequency values) is .000401. When we calculate the standard error for each of the Google corpora individually we get .00110812 for American English, .00067054 for British English, and .00095698 for German, for 1800 onwards where, according to Figure (5.6) the estimate for the variation of measurement seems most reliable.

The drop in the number of words per full stop in the second half of the 19th century would be an example of linguistic diachronic variation. Notice, however, that this has no effect on the variation about the mean. I am therefore inclined to think that for the period of 1800 onwards the variation about the mean relates to the measurement accuracy of the corpus. So if the standard error for the number of full stops per word in the corpora is .000401 and we can assume that the variation of measurement is as much as this, then that means that for every token count there is an uncertainty in the number of tokens of .000401 per token. So, for instance, to be as uncertain as one whole token, the item whose frequency we are measuring would have to have at least 2,494 tokens, in order to be unsure whether it were 2,494, 2,493, or 2,495. When we calculate this for each of the corpora individually, we get 902 tokens for American English, 1,491 tokens for British English, and 1,045 tokens for German. As long as

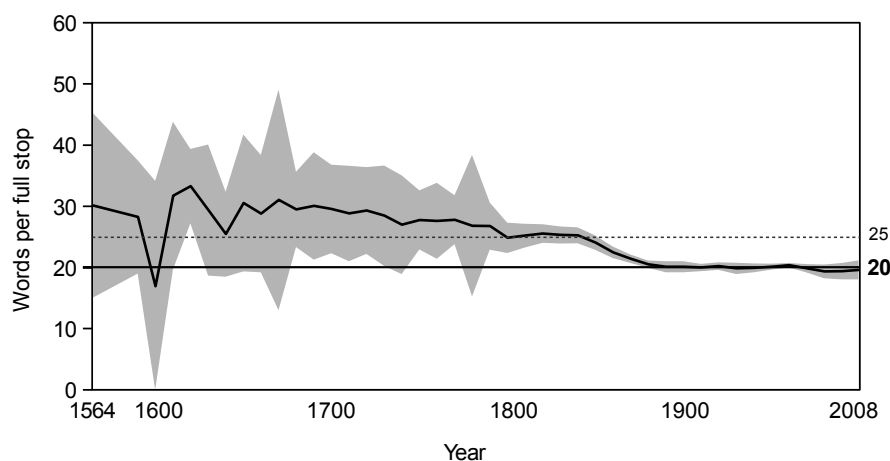


Figure 5.6: The number of words per full stop (trimmed mean, $\alpha = 95\%$) and their standard deviation as a measure of a corpus' measurement accuracy (Google Books n-gram corpora). Due to the small amount of texts written before 1800 included in the corpora, measurement accuracy is quite low ($\sigma_{\max} = 10.9$). However, after 1800 the measurement accuracy is much better. The percentage of each corpus that falls within the time interval of 1800 to 2000: 99.8% US, 99.4% English fiction (mostly AE), 96.5% UK, and 99.7% the German Google Books corpus.

the number of tokens of L1 metaphors are at least this much or greater, then the zero frequency of the novel metaphors is likely to be true. For the L1 metaphors in Google US we have 27,554 tokens, for Google UK 6,934 tokens, and 947 tokens for German. In the British and American English corpora we have thus a lot more tokens than the minimal uncertainty, in German it is almost exactly the minimum. Hence we can conclude that if we have as many or more tokens of L1 metaphors than the number of tokens required to reduce our uncertainty about the zero frequency of the novel metaphors really being zero, then the classification of the novel metaphors is most likely correct. This is most definitely the case for the British and American English corpora and most likely for the German corpora. What is important is whether participants have come into contact with these expressions, and the chances of having done so for something that is a hapax legomenon in a 300-billion word corpus are very small.

5.7 Substantive and statistical significance of differences in corpus frequency

I showed how knowledge of a corpus' measurement accuracy can help us decide whether, given the size of the corpus, searches that return zero tokens are reliable. This allowed us to determine the minimal token count necessary to constitute a reliable frequency estimate. However, this leaves open the question how to interpret frequency differences of arbitrary size and arbitrarily far away from the minimal token count given measurement accuracy. Given what we have learned about the frequency distributions of the vocabularies of American English, British English, and German, how are we to assign a measure of statistical and substantive significance to arbitrarily large frequency differences, and that with the statistical rigour available to us in experiments? Statistical significance, defined in the way it usually is in experimentation, would

have to be the improbability that a frequency difference could have been the result of chance. In the case of corpus frequencies, the null model, which we use as our model of random chance, would have to include our knowledge of measurement accuracy. Substantive significance refers to the size of an effect. If, in the statistics I set out to develop here, statistical significance is to be the chance that an observed difference in corpus frequency is larger than the corpus' measurement inaccuracy, substantive significance would have to be the difference in frequency minus the measurement inaccuracy. However, the tricky thing here is that, from what we have learned so far about vocabularies' frequency distributions, substantive significance and measurement inaccuracy must be interpreted *non-linearly*.

I propose that the same probability mass function $f_X(x) = \Pr(X = x)$ that we found for the frequency distributions should be our non-linear function to evaluate substantive significance. When frequency differences are converted to cumulative familiarity probabilities using Equation (5.6), which we had derived from the probability mass function, differences between familiarity probabilities can be interpreted like absolute differences on a linear scale. At the same time, the absolute difference between two familiarity probabilities is effectively a measure of effect size and thus substantive significance. Therefore, when we ask what the substantive significance of the difference between the corpus frequencies of two items y_1 and y_2 is, all we need to do is convert them to familiarity probabilities Y_1 and Y_2 using the cumulative probability function(s) derived from the respective corpus (or corpora) and calculate the absolute difference $|Y_1 - Y_2|$. In order to find the statistical significance of the difference, I propose the following approach.

We know that on a linear scale, whether the difference between any two points on the scale is statistically significant is limited by the precision with which the positions of the two points can be measured. We also know that because the scale is linear, this measurement accuracy has a constant value for the entire scale. As we have seen, the frequency distributions of vocabularies are non-linear. We can estimate the corpus-specific measurement accuracy by measuring the frequency variation for items that are (1) highly frequent and (2) for which we can assume that they are less likely to be subject to linguistic variation that could affect their frequency. I proposed that punctuation marks would make better candidates for such items than words, phrases or sentences, and I proposed that we take their time-differentiated standard error as an estimate of measurement accuracy. And I demonstrated how knowledge of the measurement accuracy, or rather inaccuracy, of a corpus allows us to determine how many tokens are needed *at least* before an estimate of corpus frequency becomes reliable. Equation (5.7) can be used to find the measurement inaccuracy in tokens

$$\frac{\sqrt{n_f}}{\sigma}, \tag{5.7}$$

where σ is the standard deviation of the frequency of punctuation marks and n_f is the number of frequency samples. Effectively, Equation (5.7) is the inverse of the standard error.

However, Equation (5.7) only provides us with a method to determine the probability that a frequency count is non-zero. With the limitation of measurement accuracy in mind, the statistical significance of a frequency difference is inversely proportional to the frequency of the individual items: assume we had four items, two low frequency items x_1 and x_2 , and two high frequency items y_1 and y_2 , such that their frequency differences $|X_1 - X_2|$ and $|Y_1 - Y_2|$ were equal in size. If we knew we were operating on a linear scale we could conclude that as long as the absolute differences are equal, the two differences would also have the same statisti-

cal significance, irrespective of the frequency of the items. We could determine the statistical significance by asking how likely it is that the difference was produced by measurement inaccuracy. The more unlikely it is, the more significant the differences are. On a non-linear scale we know that statistical significance is *not* independent of the frequency of the item pairs and so the likelihood of the differences being measurement inaccuracies must be mediated by a significance coefficient that changes with the frequency of the items. The rate of change in this significance coefficient is more drastic on a non-linear scale than it is on a linear one. On a linear scale we know that the same measurement accuracy applies everywhere and so the rate of change in the significance coefficient is constant. On a non-linear scale, however, the rate of change in the significance coefficient cannot be constant. Consequently, on a linear scale, the statistical significance of frequency differences decreases with the frequency of the item pairs; on a non-linear scale, statistical significance decreases more than on a linear scale because the significance coefficient changes with the frequency of the item pairs.

I propose that we take the probability mass function we derived earlier in Equation (5.4) as a candidate for the significance coefficient function of change. By assuming that the probability mass function indicates the significance coefficient, we assume that the rate of change in the coefficient is largest for frequency differences of item pairs with medium frequency, in the middle of the scale rather than on the tails. Overall, the change in the coefficient is as we would expect: the statistical significance of differences decreases with frequency. To take the probability mass function as the significance coefficient function of change is to find the cumulative familiarity probability (Equation 5.6) that corresponds to the measurement inaccuracy. We first need to convert the measurement inaccuracy we found with Equation (5.7) to its frequency equivalent. Equation (5.8) can be used to find the corpus-specific frequency in units [tokens per million wordforms]

$$\frac{n_c \sqrt{n_f}}{\sigma} \cdot 1,000,000 \quad (5.8)$$

where n_c is the size of the corpus, the number of wordforms. Then, find the probability that corresponds to this frequency using Equation (5.6). Since familiarity probabilities can be compared directly in the same way that we would on a linear scale, by converting the measurement inaccuracy to its corresponding cumulative familiarity probability we can use it directly. In other words, when both the familiarity probability differences and the measurement inaccuracy operate on the normalised scale of the cumulative probability functions that we derived from corpora, we know that the significance coefficient is constant in the cumulative distribution. Or conversely, we can think of it as using the probability mass function as our gauge of statistical significance.

With measurement inaccuracy converted to a familiarity probability, we then perform a one-sample t-test using Equation (5.9) which compares the probability difference between the two sets of metaphors Y_1 and Y_2 to the measurement inaccuracy τ . Since we are comparing across different corpora, say, cross-linguistically shared metaphors in English y_1 and German y_2 , the measurement inaccuracy is pooled from the respective corpora and expressed as a familiarity probability so that it can be compared to metaphors' familiarity probabilities. We also need the difference $|Y_1 - Y_2|$ of the familiarity probabilities of the two sample sets of metaphors, and the pooled standard error of the sample sets using their pooled standard deviation σ_p divided by their combined sample size N . When pooling the standard error, opt for the *largest* inaccuracy

estimate rather than a weighted average.

$$t = \frac{(|Y_1 - Y_2|) - \tau}{\sigma_p / \sqrt{N}} \quad (5.9)$$

Use the t distribution function or a t table to find the one-tailed p -value for the corresponding degrees of freedom. We are only interested in whether the probability difference between the two sets of metaphors is larger than the measurement inaccuracy. A one-tailed p -value is therefore enough. The degrees of freedom are one less than the total number of samples N , i.e. the number of familiarity probabilities, $N - 1$. As an alternative to a t -test, an F statistic with one degree of freedom can be used.

Find β , the probability of false negatives, as follows. First, find the critical t value that corresponds to the calculated p -value by using the inverse t distribution function and the degrees of freedom $N - 1$. Remember to double α since we have a one-tailed p -value. Then calculate the mean at the critical t using Equation (5.10)

$$\mu_{crit} = \tau + t_{crit} \cdot \frac{\sigma_p}{\sqrt{N}}. \quad (5.10)$$

Then use μ_{crit} in Equation (5.11)

$$\frac{|\mu_{crit} - (|Y_1 - Y_2|)|}{\sigma_p / \sqrt{N}}. \quad (5.11)$$

If Equation (5.11) yields a value ≥ 0 , use it in the t distribution and degrees of freedom $N - 1$ to find β . Otherwise use Equation (5.12)

$$- \left(\frac{|\mu_{crit} - (|Y_1 - Y_2|)|}{\sigma_p / \sqrt{N}} \right), \quad (5.12)$$

use the t distribution to find the corresponding probability p , then $1 - p = \beta$. With β known, we can determine the power of our statistic with $1 - \beta$. We might want to correct our estimate of the substantive significance of the difference in familiarity probabilities by subtracting the measurement inaccuracy from the difference, $|Y_1 - Y_2| - \tau$, and only report corrected substantive significances that are larger than zero. $1 - \beta$ is the statistical power. Let us summarise which of the frequency differences of pairs of metaphor types are statistically significant.

Table (5.3) summarises the substantive and statistical significance of the familiarity probability differences $|L1 - L1|$, $|S - S|$, $|N - N|$, and $|L1 - S|$. The sample sizes in Table (5.3) are the total numbers of metaphors in all groups whose frequencies we are comparing. There are 12 English and 12 German language-specific metaphors in the experiments, so $N_{|L1-L1|} = 2 \times 12 = 24$. There are 14 cross-linguistically shared metaphors in the data set, so $N_{|S-S|} = 2 \times 14 = 28$. Since there are 12 novel metaphors, we need to make sure they occur neither in English nor German corpora, so $N_{|N-N|} = 2 \times 12 = 24$. There are 12 English and 12 German language-specific metaphors. Since the familiarity probabilities of the 14 cross-linguistically shared metaphors are reasonably similar between the English and German corpora, we count each shared metaphor only as one sample, which makes it harder for the t statistic to yield a statistically significant p -value. The total sample size is thus $N_{|L1-S|} = 2 \times 12 + 14 = 38$. The estimate of statistical significance for $N_{|L1-S|} = 38$ is more conservative than if we assumed $N_{|L1-S|} = 2 \times 12 + 2 \times 14 = 52$.

Table 5.3: Corpus-based familiarity probability. Shown is the sample size N , the pooled standard error σ_p/\sqrt{N} of the two sets of metaphors being compared, the pooled average measurement inaccuracy τ expressed as cumulative probability. The effect size T is the absolute difference $|Y_1 - Y_2|$ between the two sets of metaphors Y_1 and Y_2 , Y_1 in the English corpora and Y_2 in the German corpora. Also shown is the p-value of the t statistic (***) for $p \leq .001$, the probability of Type I errors α , the probability of Type II errors β , and statistical power $P = 1 - \beta$. The last row provides general estimates of central tendency, averaged across the individual comparisons and weighted according to their sample size.

	N	σ_p/\sqrt{N}	τ	T	$p = \alpha$	β	P
$ L1 - L1 $	24	.038	.002588	.030	.243	.325	.675
$ S - S $	28	.048	.002588	.019	.370	.183	.817
$ N - N $	24	0	.002588	0	.997	0	1
$ L1 - S $	38	.036	.002588	.165	.00003 ***	.007	.993
	114					.116	.884

Of course we know that the familiarity probability of, say, the cross-linguistically shared metaphors will deviate from their mean probability, and that within the same corpus. When we then calculate probability differences and that across different corpora, the corpus-specific variances add up. Given in Table (5.3) is the pooled average standard error as a measure of variance of familiarity probabilities within the metaphor groups being compared (within-groups variation). The measurement inaccuracy pooled from the respective English and German corpora is expressed as a cumulative familiarity probability. I provide the standard error here, because it is in the same unit as the probabilities and this allows us to say, for instance, that the size of the difference in familiarity probabilities $|L1 - L1|$ of .030 is only about three quarters of the deviation of probabilities from this mean ($.030/.038 = .776$). If the variance of familiarity probabilities is as large or greater than the probability difference itself, I propose, we can assume that it is unlikely that this difference will be significant when we compare it to the corpus-specific measurement inaccuracy. The probability distribution, as I suggest, is the density function derived from the English and German corpora, and, let me emphasise, this is a new distribution and we need to investigate further what its critical values are. I am merely laying out the basic procedure here and this procedure is only as good as the *estimate* of the corpus-specific measurement accuracy. For $|L1 - S|$ we see that the probability difference $|L1 - S|$ of .165 is four and a half times greater than the combined standard error ($.165/.036 = 4.599$). We can think of the standard error as the within-groups variation; then the difference in familiarity probabilities is the between-groups variation.

An absolute difference $|Y_1 - Y_2|$ is the difference between the two sets of familiarity probabilities Y_1 and Y_2 , Y_1 in the English corpora and Y_2 in the German corpora. The absolute difference $|Y_1 - Y_2|$ can be understood as an effect size estimate T . As an effect size estimate, it is an indicator of *substantive significance*. This estimate can be corrected by subtracting the measurement inaccuracy τ from the difference, $|Y_1 - Y_2| - \tau$, in order to obtain a more realistic estimate of substantive significance. Reported in Table (5.3), however, is the uncorrected difference $|Y_1 - Y_2|$ so that the reader may check the result of the t test. $|Y_1 - Y_2| - \tau$ can be calculated from Table (5.3) as needed. The ratio between effect size T and measurement

inaccuracy τ can be understood as a measure of the resolution R of measurement, $R = T/\tau$. This gives us a similar resolution for the $|L1 - L1|$ ($.030/.002588 \approx 11.59$) and $|S - S|$ difference ($.019/.002588 \approx 7.34$). The resolution of the $|L1 - S|$ difference is about six times higher ($.165/.002588 \approx 63.76$).

For the familiarity probability differences $|L1 - L1|$ and $|S - S|$, the t statistic concludes that with a chance of $p_{|L1-L1|} = .243$ and $p_{|S-S|} = .370$, the differences could be measurement inaccuracies. For the difference $|N - N|$, the t statistic yields $p = 0$, but we know that if there is no variance and the novel metaphors can neither be found in any of the English nor any of the German corpora, then we know that the p-value should really be one, not zero (if we were measuring with infinite precision, that is). At the same time we know that because there always is measurement inaccuracy inherent in a corpus frequency estimate and this inaccuracy is larger than zero, it is unlikely yet not impossible that measurement inaccuracy would produce a measurement of zero tokens. With this in mind, the p-value of the t test statistic is uninformative. Earlier, in Section (5.6), I described how to determine the lowest *reliable* token count, with measurement inaccuracy in mind. We therefore need to keep in mind that frequency counts of absolute zero, which of course can have no variance, will produce uninformative p-values in the t statistic (or a one degree of freedom F statistic, for that matter). A p-value of one would mean we are 100% certain that $|N - N| = 0$ because the novel metaphors occur neither in any of the English nor any of the German corpora. This certainty is lessened only by measurement inaccuracy. Since Equations (5.8) and (5.6) allow us to express the measurement inaccuracy τ as a cumulative familiarity probability, a reasonable estimate of the *true* p-value might be $1 - \tau$, so $p \approx .997$. The t test finds only the difference $|L1 - S|$ to be statistically significant with a chance $p = .00003$ that measurement inaccuracy could have produced a familiarity probability difference as great as .165.

I provide estimates of statistical power, the reliability of the test statistics, in Table (5.3): the power P . In order to better understand statistical power and reliability, we need to understand the kinds of errors that can occur in the test statistics we are using to estimate the validity of hypotheses. Any test, statistical tests included, can draw two kinds of wrong conclusions: Type I and Type II errors. The two types of error and a test's likelihood to falsely yield one of them are related. As the chance of Type I errors decreases, the chance of Type II errors increases, and vice versa. In order to know the precision of a test we need to know its chance to make Type I and Type II errors. Hence, the precision of a test is its statistical power. Statistical significance is the probability of obtaining results as extreme or more extreme than those observed, given that the null hypothesis is true. Power analysis is useful in determining whether or not we might infer that the null hypothesis is likely to be true given that the test does not exhibit a significant difference, but importantly that is not generally a valid assumption.

Following Fisher (1925), the level for statistical significance is set to .05. This also means that the chance of a test statistic making a Type I error, α , to conclude there is an effect where there is none, is also 5%. Cohen (1988) assumes that researchers would view Type I errors as being four times more serious than Type II errors. So if the chance of making a Type I error is limited to no more than 5%, then the chance of making a Type II error should be no more than 20%. The power of a statistical test, a measure of its sensitivity and reliability, is the ability to detect an effect if there really is an effect to be detected. In other words, statistical power is a test's rate of true positives and the rate of true positives is $1 - \beta$, where β is the rate of false negatives. Therefore, if the chance of Type II errors, β , should be 20% or less, a statistical test

Table 5.4: Hypotheses confirmation and statistical tests.

	Hypothesis true $p(H) = \text{effect size } T$	Hypothesis false $p(\neg H) = 1 - p(H)$
Test positive	True Positives $p(D H) = 1 - \beta$	False Positives $p(D \neg H) = \alpha$
Test negative	False Negatives $p(\neg D H) = \beta$	True Negatives $p(\neg D \neg H) = 1 - \alpha$

should have a power of 80% or more.

A test's rate of true and false positives, and true and false negatives can be written as conditional probabilities. The chance that a test will correctly accept hypothesis H when there is positive evidence D supporting it can be written as $p(D|H)$, the rate of true positives. Similarly, the rate of false positives α can be expressed as $p(D|\neg H)$: the test statistic takes the evidence to be in favour of H when, really, the null hypothesis $\neg H$ is true. The rate of false negatives β is $p(\neg D|H)$: the test interprets the evidence as negative evidence, evidence in favour of the null hypothesis $\neg H$ when actually H is true. Recall that the statistical power of a test is defined as $1 - \beta$. We therefore know that $p(D|H) = 1 - \beta = 1 - p(\neg D|H)$. We can therefore write statistical power P also as $P_{p(D|H)}$. A similar tradeoff holds between the rate of false positives α and the rate of true negatives. A test's chance to truthfully reject H when the evidence is negative can be written as the conditional probability $p(\neg D|\neg H)$ and $p(\neg D|\neg H) = 1 - \alpha = 1 - p(D|\neg H)$. The relationships between these probabilities are summarised in Table (5.4).

The statistical significance of an effect, i.e. its improbability p under the null hypothesis, is an experimental estimate of a test's true α rate. In the corpus analysis, β is calculated from the t statistic using Equations (5.11) and (5.12). In the experiments of this thesis, β is determined through power analysis using G*Power (Faul, Buchner, Erdfelder & Lang, 2014). We can take effect size T as an experimental estimate of $p(H)$, the probability that the hypothesis is true, the probability that there really is a difference in reading/response times and plausibility judgements between language-specific and cross-linguistically shared metaphors. In the corpus analysis, $T = |Y_1 - Y_2|$; in the experiments, $T = R^2$. The null hypothesis $\neg H$ is the hypothesis that the effect predicted in H does not exist. The probability of H and $\neg H$ are related: as one increases, the other decreases, and the sum of the probabilities of all hypotheses must be one, $p(H_1) + p(H_2) + p(H_3) + \dots + p(H_n) = 1$. If there are only two competing hypotheses, H and $\neg H$, as in the corpus analysis, then the probability that the null hypothesis holds true is $p(\neg H) = 1 - p(H)$. Since we have experimental estimates of $p(H)$, the rate of false positives (Type I errors, $p = \alpha$), and the rate of false negatives (Type II errors, β), we can calculate the probability of the null hypothesis $p(\neg H)$, the rate of true positives $p(D|H)$, and the rate of true negatives $p(\neg D|\neg H)$. I said that in the corpus analysis, we can think of the absolute differences $|Y_1 - Y_2| - \tau$ as effect size estimates of H and τ is the variation we would expect if $\neg H$ was true. We therefore know that we can determine the probabilities $p(H)$ and $p(\neg H)$, of which we know that $p(H) + p(\neg H) = 1$, by normalising the sum of their absolute effect sizes.

If, in the corpus analysis and experiments of this thesis, we know β , we can estimate the

reliability of the test statistics which estimate substantive and statistical significance as power $P_{p(D|H)} = 1 - p(\neg D|H) = 1 - \beta$. Cohen (1988) assumes that researchers would view Type I errors as being four times more serious than Type II errors. Following Fisher (1925), I proposed to set the chance of making Type I errors α to .05. So if we take Cohen's recommendation into account, then the chance of making a Type II error should be no more than .2. The power of a statistical test is its ability to detect an effect if there really is an effect to be detected. In other words, statistical power is a test's rate of true positives. Therefore, if the probability of Type II errors, β , should be .2 or less, a statistical test should have a power of .8 or more. The β rate was calculated from the t statistic, see Equations (5.10), (5.11), and (5.12). In the analysis of the familiarity probability differences, the $|L1 - L1|$ difference has the lowest statistical power with .675, which is directly related to it having the lowest number of samples of the non-zero differences. The statistical reliability of all other familiarity probability differences is well above the value recommended by Cohen.

Given in the last row of Table(5.3) are weighted average tendencies. In the calculation of these averages, the number of samples is used as weights w_i . The weighted averages are calculated as follows. For β , the probability of Type II errors, $\bar{\beta}$ is the weighted average probability

$$\bar{\beta} = \frac{\sum w_i \beta_i}{\sum w_i}. \quad (5.13)$$

For the statistical power, \bar{P} is the weighted average power,

$$\bar{P} = \frac{\sum w_i P_i}{\sum w_i}. \quad (5.14)$$

The weighted average power \bar{P} is large with .884 according to Cohen's (1988) guidelines. We can thus be confident that the probability difference $|L1 - S|$, which the t statistic judges to be statistically significant with $p = .00003$, is reliable. We can also be confident that the differences $|L1 - L1|$ and $|S - S|$ are unlikely to be more than measurement inaccuracies ($p = .243$ and $p = .370$ respectively). We can therefore conclude with an average certainty of 88% that only the frequency difference between the L1 language-specific and the cross-linguistically shared metaphors differ significantly, but with the same certainty we know that metaphor frequencies of language-specific, cross-linguistically shared, and novel metaphors are comparable between the English and German corpora. Hence, if the frequency difference between the L1 language-specific and the cross-linguistically shared metaphors had an effect on reading/response times and plausibility judgements, it would have the same effect on native speakers of English and German, regardless of whether they are fluent in the other language or not.

5.8 Cross-linguistic comparability of metaphors' length

We intend to measure reading/response times in the experiments to come. Practised readers employ many strategies to speed up the reading process. For instance, a word can be perceived and processed in real time without being fixated (Staub & Rayner, 2007). This can also affect plausibility judgements (Staub, Rayner, Pollatsek, Hyönä & Majewski, 2007), the other experimental measure we intend to use. The number of orthographic characters perceived in one saccadic movement of the eye gives readers a time advantage. Practised readers also pay more attention to the beginning and end of a word than to the characters in the middle. Words and

sentences are generally longer in German than they are in English, even though the information communicated in the metaphors would be similar. In this section, we will therefore discuss metrics of metaphor length that may affect reading/response times and plausibility judgements. If cross-linguistic differences in metaphors' phonological, morphological, and orthographic length are unavoidable, we would want these differences to be equally distributed across metaphor types so that their effects on reading/response times and plausibility judgements are constant across experimental manipulations.

The metaphors will be presented visually in the experiments. We will therefore discuss issues of visual perception and reading strategies first. The auditory cortex, where spoken language is processed, is stimulated even when reading silently (Yao, Belin & Scheepers, 2011; Perrone-Bertolotti, Kujala, Vidal, Hamame, Ossandon, Bertrand, Minotti, Kahane, Jerbi & Lachaux, 2012; Perrone-Bertolotti, Rapin, Lachaux, Baciú & Lœvenbruck, 2014; Groenewold, Bastiaanse, Nickels, Wieling & Huiskes, 2015). We will therefore also discuss metaphors' phonological length, measured as the number of syllables. We have restricted our selection to metaphors of the form *noun is noun* or *noun is adjective*, but sometimes grammatical differences between English and German will force the presence or absence of articles (determiners). For instance, the cross-linguistically shared metaphor *ignorance is bliss* has no determiners preceding the nouns in English, but in German, *Unwissenheit ist ein Segen*, the indefinite determiner *ein* (*a*) precedes *Segen* (*bliss*). On the one hand, this causes cross-linguistic differences in metaphor length; on the other hand, the presence or absence of a definite or indefinite determiner can greatly affect the metaphoric meaning: compare *sharks are dangerous predators*, *a shark is a dangerous predator*, and *the shark is a dangerous predator*. German has a richer inflectional morphology than Modern English, which should increase metaphor length. The number and types of morphemes is therefore a fourth metric.

Practised readers employ many strategies to speed up the reading process. For instance, a word can be perceived and processed in real time without being fixated (Staub & Rayner, 2007). This can also affect plausibility judgements (Staub et al., 2007). The number of orthographic characters perceived in one saccadic movement gives a reading time advantage known as *parafoveal preview benefit*. Within the human field of view, the foveal region is the central area, 1 degree of angle on either side of the fixation (about three to four characters). This is where most attention is focused and where visual acuity is highest. The parafoveal region lies within 5 degrees of angle on either side of the fixation (15 to 20 characters). Within this region, word boundaries and text continuity information is perceived. Everything within the field of view beyond the parafoveal region is called the peripheral region, where the eye can only perceive spatial information (e.g. the end of a line). Matt Davis and his collaborators showed that practised readers do not scan over each individual character. They pay more attention to the beginning and end of a word than to the characters in the middle; they visualise the word as a whole. The following example shows the same text twice. In the first one, the beginning and end of each word remains unchanged, only word-medial characters are juxtaposed. In the second text, all characters are scrambled within a word, regardless of their original position (Matt Davis from the Cognition and Brain Sciences Unit at Cambridge University wrote an extensive blog on this citing the major developments from the last decade <http://www.mrcmbu.cam.ac.uk/people/matt.davis/cambridge/>).

Accordnig to rscheearch at Cmabrigde Uinervtisy, it deosn't mtttaer in waht oredr

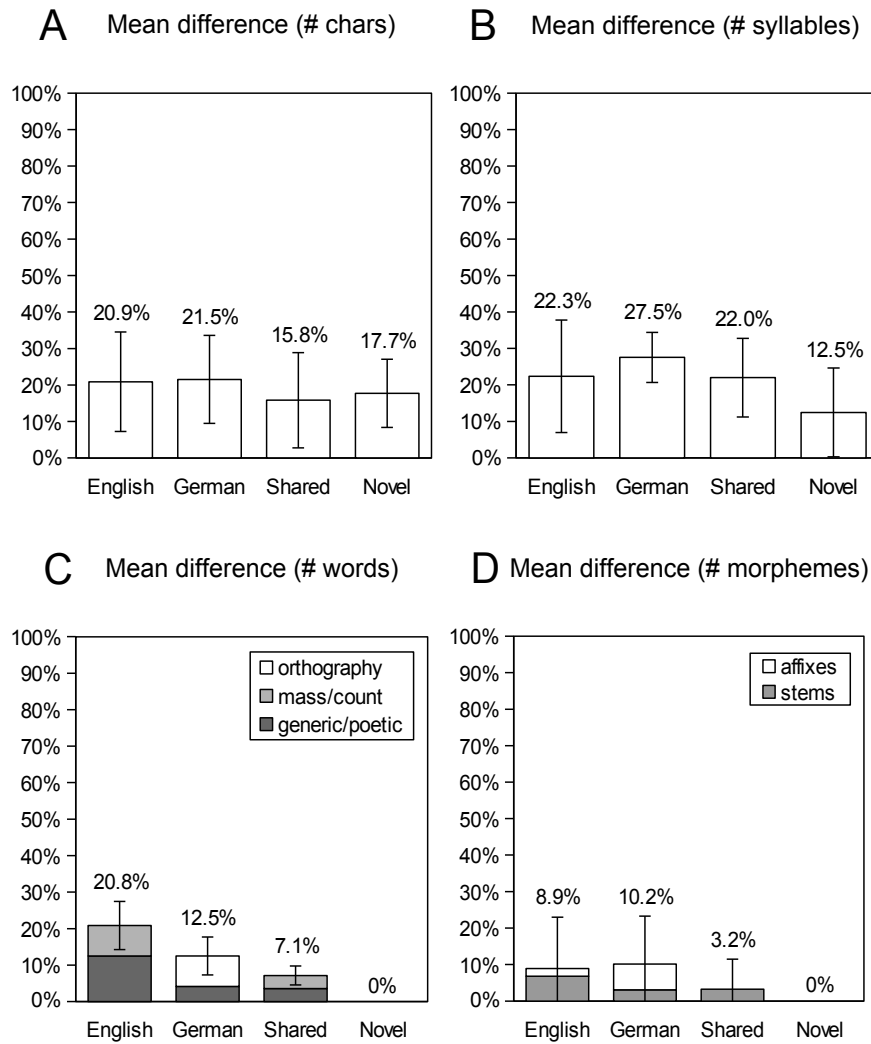


Figure 5.7: Difference in metaphor length between the English and German version of the material. Length measured as the number of orthographic characters (**Panel A**), the number of syllables (**Panel B**), the number of words (**Panel C**), and the number of morphemes (**Panel D**). The difference in the number of words is divided further into the number of cases where the difference is due to the count-noun/mass-noun distinction (e.g. the noun is countable in English but not in German or vice versa) and cases where it is due to the presence of a definite or indefinite determiner eliciting a generic or poetic reading (e.g. *love is all around* versus *the love we share*). The difference in the number of morpheme distinguishes between affixes and stem morphemes. Y-error bars show the standard deviation for each metaphor type.

the letters in a word are, the only important thing is that the first and last letters be at the right place. The rest can be a total mess and you can still read it without a problem. This is because the human mind does not read every letter by itself, but the word as a whole.

Cinocadrg to hrsecehcra ta Iamcrbegd Niurveysy, ti odesn't amrtte in twah eordr eht tletters in a word era, het yoln ipormettn ihtgn si that het trfsi nad tsla ltre eb ta hte hrgit apecl. Teh rset nac be a tlota smse and oyu nca lltsi adre it towutih a oprbelm. Siht si acbeesu hte huamn nimd deos tno dera vreye teerlt yb itsfle, tbu het wdor sa a wehlo.

The first text preserves the first and last letters that correspond to the first and last sounds (syllable onsets and codas, to be precise), while in the second text all letters within each word are scrambled, regardless of their original position. The first text is easier to read than the second text. Over the years there have been many additions to this finding which show that, as we would expect, the more the amount of scrambling increases, the harder it is to parse the text. Rayner, White, Johnson & Liversedge (2006) is one of the more recent papers that argues in favour of this gradient perspective. The important point to take away from this for our current discussion is that we, as practised readers, have a visual representation for each word in our minds in a way similar to the way that we associate the word's pronunciation with the concept of it in our mind and the object it refers to in the real world and these visual representations affect how we anticipate what we are reading, on top of orthographic rules. Nevertheless, even with the simplistic approach of counting the number of orthographic characters, we can get a first impression whether the German and English translations differ significantly. The difference in the number of characters never exceeds 21.5% and never is less than 15.8% and power analysis shows that statistical power ($1 - \beta$, the chance of false negatives, β) is .830 for English metaphors, .825 for German metaphors, .860 for cross-linguistically shared metaphors, and .856 for novel metaphors; all well above the chance level of .8 recommended by Cohen (1988). The variation in the number of characters for the translations is not significantly different between metaphor types.

Visual complexity, measured as the number of orthographic characters, factors into parsing and processing complexity (Figure 5.7, Panel A). Of course, this is a very simplistic measure for at least two reasons: Firstly, English and German have completely different ways of orthographically indicating the phonology of a word. For instance, whereas a /k/ sound in English can be written as a <c> as in *magic* /'mɑdʒɪk/, but also <ck> as in *click* /kɪk/, in German it is most commonly spelt <ck> as in *Schmuck* /ʃmʊk/ (*jewellery*). In the case of German then the number of orthographic characters is greater than in English because of the extra <c>. A <j>, on the other hand, in English can be the affricate /tʃ/ as in *jungle* /'dʒʌŋɡəl/ or *jewellery* /'dʒu:əlɪrɪ/, while in German it usually denotes the approximant /j/ as in *Junge* /'jʊŋə/ (*boy*). So when we count the orthographic character <j> the number we get in English and German might be close to the actual phonology, even though they correspond to completely different sounds in the two languages. An example where the sound denoted is the same in both English and German but spelling differences can accumulate quickly, is /ʃ/, which in English can be expressed orthographically as <sh> as in *shark* /ʃɑ:(r)k/, but in German it is always spelled <sch> as in *Schmuck* /ʃmʊk/ (*jewellery*). Secondly, in 1996, German orthography was officially changed. For instance, <ß> in many words, but not all, was changed to <ss>. For instance, *Fuß* → *Fuss* (*foot*) and *daß* → *dass* (the complementiser *that*), but not the determiner *das* (*the*) which retained just one <s> in its spelling. *Maße* /'mɑ:sə/ (*measures*, plural)

retained its old spelling to mark a phonological contrast to *Masse* /'masə/ (*mass*, singular). So depending on the age of the texts in the corpora we might get different numbers of characters for the same words because of the orthography reform and notice that this has happened within the lifetime of those participating in the experiments that follow. The mismatch between phonology and orthography, the systematic errors in counting orthographic characters, and changes in the orthography of German are one group of aspects that make the number of orthographic characters a suboptimal measure of metaphor length.

EEG and fMRI studies support speakers' experience of an "inner voice" during silent reading as neural activity in speech-selective areas of the auditory cortex (Yao et al., 2011; Perrone-Bertolotti et al., 2012, 2014; Groenewold et al., 2015). A measure that might thus seem more sensible to us could be the relative difference in the number of syllables (Figure 5.7, Panel B). The syllable is an important building block of phonology and in many ways counting syllables is a much better indicator of the pronunciation we associate with a word than orthographic features such as the number of words (recall the variable spelling of the compound nouns *pay day* and *election day*) or the number of characters (think of all the silent letters in English orthography, e.g. word-initial <k> in *know*, word-medial <gh> in *sight*, or word-final <e> in *translate*). And because of the internal structure of the syllable which includes the rhyme as important intermediate level, rhyme, as well as language-specific consonant combinations in onset and coda can be important hints that help us anticipate words and thus reading. The variation of the number of syllables is much stronger across the four metaphor types than for the number of orthographic characters, which aligns with what we would expect from the lexical-phonological differences between English and German. German-specific metaphors have the highest difference in the number of syllables with 27.5% when they are translated to English. Novel metaphors have the lowest difference with 12.5%. Power analysis shows that statistical power is .817 for English metaphors, .764 for German metaphors, .802 for cross-linguistically shared metaphors, and .892 for novel metaphors. We can therefore consider the variation in the cross-linguistic difference in the number of syllables to be equally distributed across metaphor types. As we will see later, this pattern between the four metaphor types does not mirror any of the experimental results. Notice how the English-specific and cross-linguistically shared metaphors have almost the relative difference of 22–22.3%. To correlate with the experimental results the difference for the English and German-specific metaphors should be similar and both should be different from the syllable difference of the cross-linguistically shared metaphors.

We have restricted our selection to metaphors of the form **noun is noun** or **noun is adjective**, but sometimes grammatical differences between English and German will force the presence or absence of articles (determiners). With respect to definiteness, a noun phrase can occur in three configurations: It can be a bare noun phrase, i.e. it contains only a noun and there is no (overt) determiner, neither a definite nor an indefinite determiner. And in the case of the metaphors we have also decided to disallow adjectives within the noun phrase to avoid complicating predictions of the compositional meaning of the metaphoric sentence and so the bare noun phrase really only contains the noun. The second configuration is a noun phrase with the noun and an overt definite determiner. The third configuration has an indefinite determiner in place of the overt definite determiner. There are two reasons for the presence or absence of a determiner in the noun phrase: (1) It can be triggered by and depend on whether the noun in question is a count or a mass noun. Since we are primarily interested in how well an English version of a metaphor matches a German translation and vice versa, I will not go

too much into the count-mass dichotomy, but it is important to note that it is by no means a categorical distinction in the sense that there are some nouns that are only one but never the other; if anything the two “categories” are inclusive and gradient. The naïve theoretic view would argue that count nouns always require a determiner, i.e. disallow the absence of a determiner, while mass nouns do not allow any overt determiner. In reality, however, in many cases it is possible to turn any mass noun such as *bliss* into a count noun as long as the context warrants a semantic interpretation where the “substance” that the mass noun is supposed to denote the “substance” property being the reason for disallowing definiteness or discretisation through quantification by a determiner can be quantified, i.e. thought of as something that can be divided into discrete quantities, e.g. *the love I feel for you*.

A second and somewhat related effect of an overt determiner is that it can result in a generic, more abstract interpretation that is often perceived as more poetic, e.g. *the love we have*. Again, this is a naïve, *ad hoc* classification because this effect of making the noun semantically feel less concrete and therefore more abstract can be achieved by any of the three noun phrase determiner configurations: (a) *sharks are dangerous predators*, (b) *a shark is a dangerous predator*, (c) *the shark is a dangerous predator*; arguably all three SHARK noun phrases do not refer to one individual shark or a specific group of sharks, but sharks in general, sharks as in the species. And this abstractness or generality is not the result of the noun phrase alone, but the predicate that follows, that they are dangerous predators, which is something our world knowledge tells us is a common feature of the species of sharks and this information thus reinforces our conviction that the noun phrase means to denote the species of sharks as a whole rather than individuals or an individual.

The gradience of these two supposedly categorical dichotomies could quickly turn into a problem if we were in the business of trying to clearly define these theoretical concepts, but in order to evaluate the comparability of the English and German metaphor translations, this theoretical problem has little practical relevance because regardless of what our intuition about the genericness or abstractness or substantiality is, what triggers our intuition works the same way in English and German with respect to the determiner configuration of noun phrases. Consider the shark example in German: (a) *Haie sind gefährliche Jäger*, (b) *ein Hai ist ein gefährlicher Jäger*, (c) *der Hai ist ein gefährlicher Jäger*; in (a) *sharks (Haie)* is plural, just as in English, in (b) it is singular, *Hai*, and the indefinite determiner *ein* quantifies over the noun, in (c) *Hai* is still singular, but this time the definite determiner *der* (*shark* is masculine) quantifies over it. Nevertheless, in all three configurations the shark noun phrase refers to the species of sharks rather than any particular shark or group of sharks and this is true for English as it is for German. So for our translations all we need to do in order to ensure cross-linguistic comparability is to check whether the noun phrases have the same configuration in English as they do in German.

For English-specific metaphors (Figure 5.7, Panel C) the relative difference with 20.8% is highest among the four metaphor types and this difference is partly due to a difference in the mass-count feature of the noun and partly due to the generic-poetic effect of an overt determiner. In any case the high relative difference for the English metaphors is not in itself a problem. I opted for more natural-sounding wordings in German which meant choosing a noun that was closer in meaning and frequency to its English counterpart and as a consequence of that choice that German noun will have differed with respect to the count-mass distinction or the generic-poetic effect. When we try to translate the German-specific metaphors to English we find that

the differences that are due to a difference in the presence of a determiner creating the generic reading, this too is always a consequence of choosing an English noun closer in meaning and frequency to the German noun than a real difference in how the generic reading was achieved. We also see that a much greater part of the dissonance in the translation of the German-specific metaphors comes from orthographic differences, which is merely results from the fact that *pay day* and *election day* are orthographically spelled as one word (*Zahltag* and *Wahltag* respectively). Nothing hinges on this spelling difference. As is well known English orthography is arbitrary at times and often allows multiple spellings. Arguably *pay day* might as well be spelled *payday* or *pay-day*; for *election day* we might be less inclined to spell it *electionday*, but nothing apart from habit and convention is preventing us from doing so. So if anything we might want to ignore the white area showing the orthographic difference which consequently means that the English translations of the German-specific metaphors differ by less than 3% of the number of words per metaphor. As we can see the difference for the cross-linguistically shared metaphors is again a mix of the mass-count distinction and the generic-poetic effect, but this time the difference is not a result of my own choosing. I have used the most common wordings as they are used in English and German. For novel metaphors, the number of words is exactly the same in both languages. More importantly though, if we look at the difference in the number of words across all four metaphor types, we might think it quite big given that it is 0% for the novel metaphors and 20.8% for the English-specific metaphors. Power analysis shows that statistical power is .831 for English metaphors, .892 for German metaphors, .918 for cross-linguistically shared metaphors, and .950 for novel metaphors; all well above the probability of .8 recommended by Cohen (1988). The overall variation in the number of words between the English and German translation of the material is therefore insignificant.

Instead of focusing on orthography, visual presentation, or the words' phonology, we could look at morphology as an important intermediate step in building phonological units into syntactic units (words). On top of this, if when we consider that German has much more inflectional morphology, we will want to keep track of not just how many morphemes each metaphor is composed of and whether that number is the same in its English and German version, but also what kind of morphemes they are and how many of each type in each metaphor (Figure 5.7, Panel D). The graph differentiates between affixes and stems. Stems are also called roots. The words in our case have no inflectional prefixes. English does not have inflectional prefixes anyway, but German does if you consider the circumfix (e.g. *ge-hör-t* with *-hör-* as the stem and the circumfix *ge- -t* inflecting it as the past participle) to be composed of a prefix and a suffix. Derivational affixes are ignored in the calculation because they come for free (or anyway, whichever way you want to put it) as soon as we have decided on one noun over another for a translation while inflectional morphology is something we expect to be richer in German. But this expectation is not borne out; all the affixes are suffixes indicating the nouns number and since the only two syntactic functions that noun phrases can occur in in the metaphors are subject and object, case marking is restricted to nominative and accusative. Nominative and accusative marking is diminished and English and German both mark number (only they use different morphemes to do so) and so in the end the differences are much less than we would have expected. The German-specific metaphors show the greatest morpheme differences with 10.2% because they have the largest difference in the number of affixes of all the metaphor types. The novel metaphors have no difference at all, just like we saw when we counted the number of words and of course these are directly connected. If they do not differ in the number

of words, they also do not differ in the number of stems. We might have expected them to at least differ in the number of affixes, but obviously they do not. Power analysis shows that the difference is not significant: statistical power for English metaphors is .912 (significance $p = .690$), .905 ($p = .645$) for German metaphors, .937 ($p = .880$) for cross-linguistically shared metaphors, and .950 ($p = \sim 1$) for novel metaphors; all well above the .8 chance level recommended by Cohen (1988). We see that because of the restrictions we imposed on metaphors' syntax in the selection process, the options of cross-linguistic morphological length differences due to grammar are quite limited.

This concludes the discussion of the formal aspects of the experimental material. We have looked at the simplest linguistic form of a verbal metaphor and found that it comes in one of two forms: either it is two noun phrases connected by the copula, **noun phrase A is noun phrase B**, or it is a noun phrase and a predicative adjective, **noun phrase is adjective**, connected by the copula. In both case the copula can be considered semantically vacuous so that in both cases we can say that the linguistic form makes the source and target concept of the metaphor overt as well as the conceptual mapping implied by the syntactic construction using the copula. We have seen that with the help of English and German corpora we can identify metaphoric proverbs that either only occur in English but not German, in German but not in English, in both languages, or in neither of them and we have seen that we can also use the check that the metaphors are currently frequent, as frequent in English as they are in German, that this frequency has been either stable or increasing for at least the last 30 years (the mean age of participants). And since we intend to present the metaphors visually in the experiments, those aspects that are pertinent to their visual presentation also suggest that the visual, phonological, morphological, and syntactic form of the metaphors between English and German is more similar than dissimilar. Next we look at the cross-linguistic comparability of the metaphors' linguistic meaning, specifically, their lexical semantics, the contextual relevance of cross-linguistic conceptual sense differences, and the metaphors' compositional semantics.

5.9 Summary

In this chapter I showed how large text corpora of English and German can be used to identify those metaphoric proverbs of either the form **noun phrase A is noun phrase B** or **noun phrase is predicative adjective** that can be classified into language-specific, non-specific, and novel metaphors according to their corpus frequency. At the heart of our investigation is the cross-linguistic intelligibility of language-specific metaphors. I defined cross-linguistic intelligibility in two ways: intelligibility in the strong sense requires that we compare participants plausibility judgements of language-specific metaphors with the plausibility of cross-linguistically shared, i.e. non-language-specific, metaphors. Intelligibility in the weak sense requires that we compare their plausibility judgements to those of unfamiliar, novel metaphors. I showed how we can ensure that the classification of the metaphors into the four types English-specific, German-specific, cross-linguistically shared, and novel can be based on synchronic and diachronic corpus analyses. I demonstrated a method that aims to derive an estimate of metaphors' likelihood to be familiar to participants as a means to compare corpus frequencies cross-linguistically and I showed how reliable the large corpora are for a diachronic frequency analysis for at the last 30 years, the mean age of participants, and up to 200 years for many of them. Based on this, we can be quite certain that the language-specific metaphors are as frequent, and thus likely

to be familiar to participants, in English as they are in German and the cross-linguistically shared metaphors are also as frequent in English as they are in German. And I showed how an assessment of the measurement accuracy of the largest corpora allows us to also conclude that the novel metaphors are truly novel and not highly infrequent language-specific or cross-linguistically shared metaphors. Similarly, these methods can be reliably used to ensure that the words that form metaphors' source and target concepts are comparable between the English and German translation of the experimental material as well. I argued that the linguistic form of the metaphors and the familiarity of that form is much more similar between the English and German version of the material than dissimilar. I have also made an effort to suggest ways to quantify these different aspects of linguistic form so that we can look for systematic variation within and between metaphor types that may explain participants' ratings in the experiments that follow. Ideally we want their ratings to only depend on their knowledge of metaphors' idiomatic meanings, but quantifying the different measures of cross-linguistic comparability of the material allows us to anticipate and quantify their influence.

Chapter 6

Material: cross-linguistic comparability of metaphors' linguistic meaning

6.1 Introduction

At the heart of the experiments that follow is the question of how speakers comprehend metaphors from a language they are not familiar with. Native speakers of that language will know idiomatic meanings for these metaphors, but speakers of another language will not. In an ideal world we would therefore want the metaphor material in the experiments to differ only with respect to these idiomatic meanings, while we would want the linguistic form of the metaphors, their lexical, and compositional semantics to be identical between the two languages we are comparing, English and German. In the previous chapter we have looked at the linguistic form of the metaphors and overall we can say that their form is much more comparable than it is different. In this chapter we look at the linguistic meaning of the metaphor material which consists of four types of metaphors: cross-linguistically shared metaphors, English-specific, German-specific, and novel metaphors.

For cross-linguistically shared metaphors there are well-established conventional wordings in the two languages, but we need to make sure that they are conceptually similar. If, according to CMT, speakers use basic conceptual metaphors to construct complex linguistic metaphors in their minds, we need to make sure that English speakers and German speakers are using the same conceptual metaphors. For language-specific metaphors, we need to ensure that, although native speakers will not have idiomatic meanings for the L2 metaphors, those language-specific metaphors not from their native but the other language, they are likely to rely on mental concepts and analogies that L2 speakers would have. For novel metaphors, none of the speakers have any idiomatic meanings and so we need to ensure that whatever is conceptually feasible for a metaphor's intended meaning, would be equally feasible for speakers of either language.

In an effort to provide a quantitative measure of the cross-linguistic comparability of the metaphors in terms of their linguistic meaning, Sections (6.2) and (6.3) describes a method adapted from computer science called Levenshtein distance. We take a look at the lexical and compositional semantics of the metaphor material and I pay special attention to each metaphor

source and target concept and whether there are senses that differ between English and German and when they do, whether the differing senses could have effects on the metaphoric mapping relevant in the context of the metaphors they appear in: for instance, TIME in English can, among other things, refer to metre in music while this is not true for German; the concept TIME thus has a sense in English that the equivalent German concept ZEIT does not have. The method adapted from Levenshtein distance allows us to quantify the number of senses that would have to be changed to go from the English meaning of a given concept to its German counterpart.

I will argue that the linguistic meaning of the metaphors is cross-linguistically much more similar than different. At the end of this chapter I will summarise the quantitative cross-linguistic comparability of the material from this and the previous chapter. The reason is that when we look at the experimental results, we can return to this summary and see if participants' rating behaviour in the plausibility judgement task can be attributed to systematic differences in the metaphors' linguistic form, familiarity, and meaning. To this end the table allows us to look at the cross-linguistic comparability of the material for each metaphor type individually so that we may see if disparities in the English and German version of the material occur more for, say, English metaphors than cross-linguistically shared metaphors. Maybe the differences in form and meaning are small on their own so that none of them individually would be the cause. But what if they added up for one metaphor type more than for another? In order to check this the table also contains a measure of the accumulated differences within and between metaphor types.

6.2 The lexical semantics of metaphors' source and target concepts

All metaphors in the experiments are either of the form **noun A is noun B** or **noun is adjective**. The copula can be considered semantically vacuous and therefore contributes nothing to the compositional meaning of the entire clause. Thus the compositional meaning of the clause is the interaction between the lexical semantics of the metaphor source/vehicle and target/tenor and the contribution made by pragmatic enrichment which is the subject of the experiments. We should therefore aim to minimise the sense discrepancies in each English-German translation pair of metaphor source/vehicle and target/tenor. For the lexical meanings of the English words, all sense comparisons are based on the Oxford English Dictionary and WordNet. For German they are based on Duden, the official German dictionary.

Two thirds of the words have more than 80% of their senses in common and of the remaining 20% that they might not have in common, none are relevant in the context of the metaphors. For instance, *time* in English can refer to metre in music, but its German counterpart in the translation, *Zeit*, cannot be used in this way. However, the sense of *time* as musical metre is not relevant in the cross-linguistically shared metaphor *time is money* and so we need to treat it differently from sense differences that are relevant in the context of the metaphors. For the remaining third of translation word-pairs the semantic overlap gradually goes down, but never lower than 20% and even if the ratio of senses that the English and German words in a pair have in common is that small, the senses that are different are more than 85% of the time irrelevant in the context of the metaphors. Even though there are slight differences in the number of

senses per translation word-pair between the metaphor categories, analysis of variance shows them to be insignificant ($F(3, 75) = 0.781, p = 0.508$) and in cases where differing senses are relevant to the comprehension of the metaphor, variation between the metaphor categories is evenly distributed ($F(3, 75) = 0.439, p = 0.726$).

We can quantify *how similar* the meaning of the English and German translations of the material are by asking how many senses would we have to add, subtract, or swap to convert the lexical meaning of an English word into its German counterpart. Levenshtein distance is a metric in information theory for measuring the differences between two strings of symbols by counting the minimum number of insertions, deletions, and substitutions of symbols it takes to convert the first string into the second string. When we use word senses as symbols in the Levenshtein algorithm, the Levenshtein distance matches our rationale for the cross-linguistic similarity of the senses of the English and German words in the metaphors and is thus a metric to quantify that similarity. For example the Levenshtein distance between *kitten* and *sitting* is 3:

1. kitten → sitten (substitution of “s” for “k”)
2. sitten → sittin (substitution of “i” for “e”)
3. sittin_ → sitting (insertion of “g” at the end)

The Levenshtein distance between strings a and b is given by $\text{lev}_{a,b}(|a|, |b|)$

$$\text{lev}_{a,b}(i, j) = \begin{cases} \max(i, j) & , \min(i, j) = 0 \\ \min \begin{cases} \text{lev}_{a,b}(i-1, j) + 1 \\ \text{lev}_{a,b}(i, j-1) + 1 \\ \text{lev}_{a,b}(i-1, j-1) + [a_i \neq b_j] \end{cases} & , \text{else} \end{cases} \quad (6.1)$$

where $\max(i, j)$ corresponds to the condition in which the two strings are of different length and have no characters in common. $\text{Min}(i, j) = 0$ corresponds to the condition in which the two strings are identical. If they are not identical, there are three options given in the minimum: the first element in the minimum corresponds to deletion, the second to insertion, and the third to match or mismatch, depending on whether the respective characters are the same.

The Levenshtein distance is zero if and only if the two strings are identical; otherwise the lower bound, the smallest distance, is at least the difference in character length between the two strings (e.g. *spoon-spoons*: Levenshtein distance = 1, insertion of “s”) and the upper bound, the largest distance, is at most the length of the longer string (two words that are of different length and do not have any characters in common, e.g. *have-is*: Levenshtein distance = 4, 2 substitutions + 2 insertions). Knowing these generalisations allows us to compare across pairs of strings where word length differs greatly. In other words, the degree of possible differences increases with the length of the words in a pair; knowing the lower and upper bounds of possible Levenshtein distances for each pair allows us to normalise the actual distances relative to these bounds so that we can compare across pairs.

Similarly, we can define the semantic distance as the number of insertions, deletions, and substitutions of senses it takes to convert one concepts’s semantics into another concept’s semantics. Some concepts only have one sense, for instance INFINITY, but most concepts have multiple senses, for example TIME: the continuum from the past through the present unto

the future, a period of time including such periods of special social importance such as the duration of political office, prison sentence, or sporting event, or a specific point or event in time, a particular historic period of social significance, the time of day, the time as given by a clock, the fourth dimension of spacetime as defined in physics, and metre in music. For any two concepts we can determine how many senses would have to be inserted, deleted, or substituted to go from the English concept to its German counterpart. Hence if we want to compare how similar the English concept of TIME is to the German concept ZEIT (time), we count how many senses they each have. If they have the same number of senses, they might still have different senses and so we count how many senses are different, i.e. how many senses would have to be substituted. If the number of senses differs between them, then in addition to counting the number of different senses we also need to count the number of senses we would have to add or remove. Of course, this metric does not tell us how far removed the two mismatching senses are in terms of speakers' intuitions, but it is a first step towards an objective measure that allows us to evaluate the experimental material shown to English and German speakers because in order to conclude whether English and German native speakers mirror each other's behaviour in the experiments, we need to determine whether the material used is comparable on formal and semantic grounds. To compute the semantic distance we already need to employ our intuition to determine when two senses of two different concepts are the same or not. Sometimes this decision seems easy and trivial. For instance, the concept TIME can refer to a stretch of time and a single occasion or event in time and these two senses exist in both English and German. Hence it is quite easy to determine that these two senses are the same for the English and German concept TIME. And it is similarly easy to determine that only in English TIME can be used in the sense of metre in music, i.e. rhythm as given by division into parts of equal duration, while this sense is not available for the concept ZEIT (time) in German (but is a sense of a different concept, TAKT (beat), i.e. is expressed by another word). However, this decision might be more difficult in other cases, such as English *bliss* and German *Segen*, which literally is *blessing*. I think it more useful to discuss these mismatches for the metaphors as a whole because in a lot of cases the senses that differ are not relevant to the properties that are transferred between the metaphor source and target concepts and so mismatches that seem important on the lexical level might not be important on the level of the metaphoric meaning. Let me therefore end this section with Table (6.1) which summarises the relative Levenshtein sense differences between the English concepts and their German counterparts and refer you to Table (A.5) in Appendix A for a complete list of these differences.

When we look at the results in Table 6.1, we see that (1) the majority (66%) of word translation pairs have very much the same meanings in English and German with a relative Levenshtein distance of less than .2, i.e. they differ in less than 20% of their senses, (2) the distribution of Levenshtein distances within each type of metaphor is very similar to the overall distribution (column %), (3) the distribution of distances between the types of metaphor is also very similar ($F(3, 75) = .781, p = .508$), (4), although differing senses become more important with increasing Levenshtein distance (%Rel), within each type of metaphor it is low and comparable across metaphor types ($F(3, 75) = .439, p = .726$).

Table 6.1: Relative Levenshtein distance of lexical semantics. The first column gives the relative Levenshtein distance (Lev) from 0 (the English and German word have the same number and kind of senses) to 1 (the English and German word have completely different senses), the second column gives the percentage (%) of words that fall within a distance band (lower limit < Lev ≤ upper limit), and whenever the meaning of the English and German word are different, the third column (% Rel) gives the percentage of how many of the differing senses are relevant in the context of the respective metaphor. Columns four through seven give the percentage of words that fall within a distance band (same as column % Rel) but for each metaphor type individually: English-specific (E), German-specific (G), cross-linguistically shared (S), and novel (N); in brackets is the % Rel for that metaphor type.

Lev	%	% Rel	% E (14)	% G (11)	% S (14)	% N (0)
0	60	0	60	56	61	59
0.2	6.7	0	15	4	11	11
0.4	11.7	14.3	8	12	12	12
0.6	13.3	5.5	17	12	12	18
0.8	8.3	23.1	0	16	4	0
1	0	0	0	0	0	0

6.3 Metaphors’ compositional semantics and idiomatic meanings

L1 metaphors have no idiomatic meanings in the L2. For L1 metaphors we thus need to ensure that the lexical meanings of the metaphor source and target in the L2 are as close to the lexical meanings of their source and target counterparts in the L1 as possible. The objective of the experiments is to find out if this is enough for L2 monolinguals to infer metaphoric meanings similar to the idiomatic meanings L1 monolinguals have—either without or with the help of additional context. The copula can be considered semantically vacuous in both English and German and therefore contributes nothing to the compositional meaning of the entire metaphor. For the novel metaphors, which obviously have no idiomatic meanings in either language, we should strive for a similar comparability of sources and targets’ compositional lexical semantics—despite the fact that there are no idiomatic meanings for monolinguals of either language with which to compare the closeness of these compositional meanings. Different from the language-specific metaphors, for shared metaphors we need to ensure that they do have the same meanings in the L1 and the L2. In each a–b pair of examples below, example (a) is the English wording of the cross-linguistically shared metaphor and (b) is the German one. Twelve of the fourteen cross-linguistically shared metaphors use exactly the same metaphor sources and targets in English and German and, as we will see in the discussion of examples that follows, also have exactly the same idiomatic meanings (including connotations) in English and German. As in the previous section, all sense comparisons are based on the Oxford English Dictionary and WordNet for English and the Duden, the official German dictionary. Throughout the discussion of the examples that follow, whenever I note a difference in sense, I have counted this difference toward the summary in Tables (6.1), (6.2), and (6.3)—even when I have noted that they might not be relevant in the context of the respective metaphors, thereby giving us a conservative quantitative estimate.

Cross-linguistically shared metaphors

(1a) Life is a journey.

(1b) *Das Leben ist eine Reise.*
The life is a journey.
'Life is a journey.'

Life has exactly the same senses in English as *Leben* in German. They range from the general experience of being alive to specific social states of that experience, the thrill of feeling alive, a particular art and style of living, or the purpose and meaning of one's life, the biography of a particular individual, the mundane everyday existence, or a liveliness in action or expression in the most general sense. *Reise* has the same meaning as *journey*, the act of travelling from one place to another, especially over long distances and following a specific plan. In German it can also be used as jargon for *trip*, a drug-induced hallucinatory experience which gives it connotations similar to the semantic field that *journey* belongs to in English because *trip* can also be used synonymously to *journey*.

(2a) Time is money.

(2b) *Zeit ist Geld.*
Time is money.
'Time is money.'

Most of the senses of *Zeit* are the same as the senses of *time* in English, the continuum from the past through the present unto the future, a period of time including such periods of special social importance such as the duration of political office, prison sentence, or sporting event, or a specific point or event in time, particular historic period of social significance, the time of day. However, in English *time* can refer to metre in music, which is not a sense of *Zeit* in German (it is a sense of the German word *Takt* instead). Notice that *time* in the sense of metre is relevant in the context of this particular metaphor because it is ruled out by the idiomatic metaphoric meaning that both English and German speakers will be familiar with. *Money* has exactly the same meaning in English as *Geld* in German. Its senses range from a mode of currency in the most general fiscal sense to the coins and paper used as physical currency to modern-day virtual credit as well as all senses of *money* relating to its social status.

(3a) Seeing is believing.

(3b) *Sehen ist Glauben.*
Seeing is believing.
'Seeing is believing.'

Seeing means exactly the same thing in English as *sehen* in German, the act of observing with one's own eyes, especially when this act of perceiving gives rise to some change of one's mental state as the result of sudden realisation. *Believing* and *Glauben* have exactly the same senses in English and German, the cognitive process that leads to convictions, especially if those convictions are not based on fact or if the process takes place in a religious context.

(4a) Love is blind.

- (4b) *Liebe ist blind.*
Love is blind.
'Love is blind.'

Love means the same thing in English as *Liebe* in German. The experiencing of strong positive emotions of regard and affection for someone, especially a romantic lover, family members, or friends, it can refer to the person that is the object of that devotion, or the act of physical love, and passion. The only difference is that in English it can refer to a score of zero in tennis or squash which is not a sense of *Liebe* in German, but as with *time* in the sense of metre in music, this sense of *love* is not relevant in the context of the metaphor. *Blind* has the same range of senses in English and German. Its senses range from the medical condition of lacking the sense of eyesight to the mental state of wilful or unintended ignorance to facts, events, or circumstances.

- (5a) Jealousy is poison.
(5b) *Eifersucht ist Gift.*
Jealousy is poison.
'Jealousy is poison.'

Jealousy and *Eifersucht* in both languages can mean a feeling of envy, but unlike the English *jealousy*, *Eifersucht* cannot be used to mean zealous vigilance as in *cherish their official political freedom with fierce jealousy*—Paul Blanshard, but, of course, German speakers would expect this behaviour from someone experiencing jealousy and in that respect they are comparable. *Poison* and *Gift* have exactly the same range of senses in English and German, from something or someone that is a harmful or destructive to a toxic substance that causes physical pain or even death.

- (6a) Revenge is sweet.
(6b) *Rache ist süß.*
Revenge is sweet.
'Revenge is sweet.'

Revenge and *Rache* have exactly the same meaning, an action taken in return for an injury or offence and it activates the same semantic field of related concepts such as *retaliation*. *Sweet* and *süß* share the same range of senses, from one of the five primary tastes that is excited by sugar present in food and is generally considered pleasant, to anything that is pleasant, especially with regards to a person's character.

- (7a) The mind is a sponge.
(7b) *Der Verstand ist ein Schwamm.*
The mind is a sponge.
'The mind is a sponge.'

Mind and German *Verstand* have almost the same meaning. The senses they share include the centre of the psyche, of judgement, intellect, and reason, it is used in connection with people who excel at reasoning, it is also common in other idiomatic expressions shared between the two languages, such as *losing one's mind/den Verstand verlieren* (identical literal form) or *being*

mindful/mit Verstand an eine Sache herangehen (literally: *to approach a thing, meaning issue here, with one's mind*). Unlike English *mind*, *Verstand* cannot be used in the sense of recall or remembrance, as in *it came to mind*; this sense is part of another word in German, *Sinn* (*sense* or *wits*). This difference could be relevant in the context of the metaphor since a *sponge* has holes and the metaphor thus alludes to the mind having holes, i.e. being forgetful; but we can also see how German speakers may derive this inference as an *ad hoc* concept given the senses of *Verstand*. English *sponge* and German *Schwamm* have the following senses in common: it can refer to a primitive multicellular marine animal whose porous body is supported by a fibrous skeletal framework or it can refer to any artificial fibre with a similar structure, in both languages all common senses focus on the feature that a sponge because of its porous fibre structure is able to soak up liquids and in both languages it can therefore be extended to anything or anyone who is quick to absorb something, including knowledge. The only sense difference is that in Bavarian, Austrian, and Swiss dialects of German, *Schwamm* can also be used to refer to mushrooms, but not that this is because they share a similar porous fibre structure with *sponges*, i.e. the focus is still on the feature relevant in the metaphor.

(8a) The brain is a computer.

(8b) *Das Gehirn ist ein Computer.*
 The brain is a computer.
 'The brain is a computer.'

English *brain* and German *Gehirn* share the sense that it is the centre of the nervous system located in the head and is colloquially associated with a person's intellect and body of knowledge. However, in English, to be the *head of the operation* is not the same as being the *brains of the operation*. *Head of the operation* is metaphoric, and not just a use of a polysemous sense of *head*. Being the *head of the operation* is not interchangeable with the other idiom. *He was the head of the operation*, but not *he was the brains of the operation*, has a literal match in the German idiom *er war der Kopf der Bande* (literally *he was the head of the band (of robbers)*). *Computer* has the same meaning in English and German. A common synonym used for *computer* in German is *Rechner* (literally *computer, calculator*), thereby activating the same semantic field of calculating (note *Rechner* is never taken to be a short form of *Taschenrechner*, literally *pocket calculator*, although they belong to the same semantic field).

(9a) A job is a jail.

(9b) *Ein Beruf ist ein Gefängnis.*
 A job is a jail.
 'A job is a jail.'

Job and *Beruf* have the largest difference in their senses out of all of the cross-linguistically shared metaphors. In German *Beruf* refers exclusively to one's occupation or profession whereas in English *job* can in addition refer to any task or action undertaken. Note, however, that the word *job* has been borrowed into German recently where it is now used as an alternative to *Beruf* whenever it refers to an undertaking that is not considered a traditional and/or permanent occupation, e.g. freelancing or helping around the house as a favour, and in that sense comes closer to the English meaning which expands the semantic field activated by *Beruf* for German speakers. In German *Gefängnis* is used indiscriminately for prisons and jails and so it includes all senses of *jail*.

(10a) Patience is a virtue.

(10b) *Geduld ist eine Tugend.*
Patience is a virtue.
'Patience is a virtue.'

Patience/Geduld and *virtue/Tugend* have exactly the same meanings in English in German, from their general role in morality, both in a Christian and Enlightenment tradition, including the archaic use of *virtue/Tugend* in the sense of *chastity* in both English and German, to the general sense of *patience/Geduld* outside its specific moral sense.

(11a) Knowledge is power.

(11b) *Wissen ist Macht.*
Knowledge is power.
'Knowledge is power.'

English *knowledge* and German *Wissen* have exactly the same meaning and activate the same semantic field as *mind/Verstand* and *brain/Gehirn*. German *Macht* has the same meaning as English *power* and activates the same semantic field in both languages relating to physical strength of a person or in an engineering context, political strength and power, military, and socio-economic strength and power.

(12a) Anger is a volcano.

(12b) *Wut ist ein Vulkan.*
Anger is a volcano.
'Anger is a volcano.'

In both English *anger* and German *Wut* share the sense of a strong emotion directed toward some real or supposed grievance or wrong-doing, but in German it may also refer to *rabies*. *volcano* and *Vulkan* mean exactly the same thing in both languages, a mountain formed by volcanic material that erupts as the result of seismic tensions underneath. In both languages the focus on the cause of the eruption of a volcano allows the metaphor to transfer this property unto emotions.

The last two, (13a)–(13b) and (14a)–(14b), unlike the other cross-linguistically shared metaphors, differ in the choice of words, i.e. their linguistic form, but I will argue that they, too, have the same idiomatic meaning in English and German.

(13a) Ignorance is bliss.

(13b) *Unwissenheit ist ein Segen.*
Ignorance is a blessing.
'Ignorance is bliss.'

English *ignorance* and German *Unwissenheit* (literally *non-knowledgibility*) mean the same thing, the lack of knowledge or education. Although the terms *bliss* (13a) and *blessing* (13b) are rooted in religion, the English and German wording of the metaphor can also be applied to any profane happenstance of good fortune. The concept BLISS (13a) describes the feeling of extreme happiness, joy, or ecstasy and in the religious sense that joy comes from the belief

in salvation. The concept BLESSING (13b), on the other hand, is the granting of a favour or the fulfilment of a wish and in the religious sense the favour granted is God’s protection, in life and beyond, which in turn includes the promise of salvation for the true believer and that is how the two concepts are connected: granting a favour through a blessing *is* bliss and so (13a) and (13b) have the same idiomatic meaning and both languages allow using it in a religious or non-religious sense.

(14a) Silence is golden (Adj).

(14b) *Schweigen ist Gold.*
 Silence is gold (N).
 ‘Silence is golden.’

There is a difference in perspective between English *silence* and German *Schweigen*. In English, *silence* refers to both the absence of sound as well as the refusal (muteness) or inability to speak. *Schweigen* only has the second sense of *silence*, the refusal or inability to speak because the first sense is part of the meaning of the word *Stille* (*silence*, literally *stillness (of sound)*). Notice, however, that *Schweigen* and *Stille* belong to the same semantic field and they both may focus on the feature of speech when appearing in the relevant context. English (14a) uses the adjective *golden* while German (14b) uses the noun *gold*, both of which conceptually mean “made of gold” or “having the appearance and/or properties of gold” and in this specific metaphor these properties are not the physical appearance, but the cultural knowledge that gold has served as a currency—either directly as the material that coins are made of or as countervalue for the trading value of currency—e.g. the word *gulden* for the currency and type of coin derives directly from the term *golden*. Gold is valuable and wealth is often measured in a currency whose countervalue is weighed in gold and wealth is often associated with a person’s fortune in life. Hence the analogy implied by the metaphor invites us to see that it can be valuable to keep silent, that it can be fortuitous to not say anything. And this cultural knowledge is as true for speakers of English as it is for speakers of German and so the idiomatic meaning of the metaphor is the same in both languages. Keep in mind that these are the conventionally accepted linguistic forms of the idiomatic metaphoric meaning in the two languages and that these are well-established historically (see Figure 5.4), so we may conclude that whatever sense differences there are on the lexical level, they are overwritten on the idiomatic level.

English-specific metaphors

(15a) Talk is cheap.

(15b) *Reden ist billig.*
 Talk is cheap.
 ‘Talk is cheap.’

Talk and *Reden* have exactly the same meaning in both languages. In both English and German *cheap* and *billig* can be used in the following two senses: (1) low in cost, easily affordable, and (2) of low quality or virtue if attributed to a person. This means they share all senses as well as connotations. Hence the compositional semantics of the German translation should be the same as in the English original. The proverb *talk is cheap* contrasts with the cross-linguistically shared metaphor *silence is golden*: *silence is golden* only forms one half of a larger proverb

which reads *speech is silver (but/and) silence is golden*. *Speech is silver* affirms that SILVER, and thus SPEECH, is valuable. *Talk is cheap* affirms that TALK can be offered at no cost and therefore lacks the value that deeds would possess.

(16a) Love is a journey.

(16b) *Liebe ist eine Reise*.

Love is a journey.

‘Love is a journey.’

I have already pointed out for the cross-linguistically shared metaphor *life is a journey* that *journey* and *Reise* have exactly the same meaning and that in German *Reise* activates the same semantic field including the drug-related sense of English *trip*, which in English belongs to the same semantic field as *journey*. Also, this metaphor might be intelligible to German speakers if they draw a conceptual analogy to the cross-linguistically shared metaphor *life is a journey*: *love* is an integral part of life, hence, if life is a journey, love might also be considered a journey since we can expect that speakers of German have similar cultural ideas about the stages of a relationship. Also recall that for the cross-linguistically shared metaphor *love is blind*, I pointed out that the only difference between the English and German senses of the word *love* (*Liebe*) is that in English it can refer to a score of zero in tennis or squash which is not a sense of *Liebe* in German, but I also said that this sense is not relevant in the context of the metaphor.

(17a) Words are daggers.

(17b) *Worte sind Dolche*.

Words are daggers.

‘Words are daggers.’

The concept *word* in English shares several senses with the German *Wort*: both mean the smallest linguistic unit with sound, content, and meaning, the spoken word, lyrics or texts (*she put her thoughts into words, he wrote both words and music*), a speech or an utterance of someone that is considered of special importance or that is quoted (*in her own words*), it can mean an angry dispute (*I had a strong word with him earlier*), or metonymically for someone’s linguistic ability (*she has a way with words*). However, in German it cannot be used to refer to an actor’s lines while at the same time it has a specific meaning in the field of theology. In English, *word* also has a theological sense, as in *In the beginning was the Word*, for instance, or *the Word made flesh*. *Dagger* in English has exactly the same meaning as *Dolche* in German, a short knife with a pointed often double sided blade. *Dolch* is a better choice for the translation than German *Messer* (*knife*) because *Messer*, although it is the more common term in modern German, *Dolch* bears a more archaic meaning that is more faithful to the meaning of *dagger*, especially because it is often thought to imply that the knife is used in a ritualistic way or for a special purpose rather than as an everyday cutting tool. Precisely because *Messer* is the more generic term here, it is less appropriate in the translation. And because of its archaic sense *Dolch* is associated more with poetic and lyric language, similar to the register and style that the English metaphor has for English native speakers.

The corpus analysis found no instances of the proverb *The pen is mightier than the sword* in German, but that is reminiscent of another metaphor that talks about *language* and *weaponry*: the proverb *the pen is mightier than the sword* is a cross-linguistically shared metaphor with a

close translation in German as *Die Feder ist mächtiger als das Schwert* (*Feder* literally means *feather*, referring to the quill used for writing, i.e. closer to the historic meaning of *pen*, it also occurs in the compound nouns *Federkiel*, *feather quill*, and *Schreibfeder*, literally *writing feather*; the rest of the proverb is a word-for-word translation of the English version). *Pen* (*Feder*) is a metonym for writing, but the way that the proverb is used, it can also be said to refer to language in general, spoken or written. *Sword* (*Schwert*) is a metonym of warfare or physical dominance, when the proverb is taken in its most general sense. Notice the conceptual similarity of the imagery. The proverb *the pen is mightier than the sword* is commonly thought to refer specifically to a political context, but it can also be read in a general sense to mean that what someone says can often be more forceful, powerful, and have a greater impact on those addressed than physical strength, power, or even violence. With this in mind we can read *words are daggers* as a consequence that follows when this general statement is taken to be true: what someone says can have a negative impact on those addressed, it can be hurtful, so if *language* is like *physical action* in its impact on people, when *language* has a negative impact on people it is also as if physical pain and injury was inflicted on them and so in that case *language* has the same effect psychologically that a weapon would have physically. Since native speakers of German are familiar with the cross-linguistically shared proverb *the pen is mightier than the sword* they might use this conceptual analogy which would make the English metaphor *words are daggers* intelligible to them.

(18a) Love is a battlefield.

(18b) *Die Liebe ist ein Schlachtfeld.*
 The love is a battlefield.
 ‘Love is a battlefield.’

This metaphor found its way into English recently through the lyrics of the song written by Holly Knight and Mike Chapman, performed by Pat Benatar, which was released in 1983 and was in the Billboard charts for 4 weeks and peaked the Hot 100 charts at #1 and the Top Tracks charts at #5 in the US and also reached #1 in Australia and Europe. It has become one of the most iconic and recognisable songs of the 1980s. This song was also very popular in Germany and so German native speakers might be familiar with the original English phrase, however, the corpus analysis shows that it has not been established in its translation. Recall that for the cross-linguistically shared metaphor *love is blind* and the English metaphor *love is a journey* I said that the only sense difference between English *love* and German *Liebe* is that in English it can refer to a score of zero in tennis or squash which is not a sense of *Liebe* in German, but I also pointed out that this sense of *love* is not relevant in the context of the metaphor. The noun *battlefield* (*Schlachtfeld*) is a compound noun in both languages. In English it consists of the two nouns *battle* and *field* which literally translate to the two nouns in the German compound noun, *Schlacht* (*battle*) and *Feld* (*field*), respectively. The individual nouns as well as the compound noun as a whole have precisely the same meanings in English and in German.

(19a) Revenge is a bitch.

(19b) *Rache ist eine Zicke.*
 Revenge is a young goat.
 ‘Revenge is a bitch.’

I have already noted for the cross-linguistically shared metaphor *revenge is sweet* that English *revenge* and German *Rache* have exactly the same meaning. A literal translation of *bitch* would have been *Hündin* (a female dog), but *Hündin* does not have the same sense as *bitch* in its profane uses. *Zicke* (a young goat) is the German equivalent of English *bitch*, used as a profanity in both languages, and it is the idiomatic meaning of *bitch* in the profanity that the English metaphor focusses on, not *bitch* in the sense of a female dog; although the profanity derives from this original sense, as in the phrase *son of a bitch*. That *bitch* and *Zicke* are equivalent profanities can be seen by its use as a noun and verb in both languages, e.g. in *bitch around*, *herumzicken*. The prefix *herum-* literally means *around* in German, hence we can consider them to be conceptually similar apart from the different mental image of a female dog versus a young goat.

(20a) A home is a castle.

(20b) *Ein Zuhause ist eine Burg.*
 A home is a castle.
 ‘A home is a castle.’

Zuhause means home, the place where someone lives, in the context of the metaphor, but note that German uses another, related Germanic word, *Heim*, to refer to other kinds of homes, such as *nursing home*, *Pflegeheim* (*Pflege* means nursing), and *retirement home*, *Altersheim* (literally *home for the elderly*). The use of the word *Heim* to mean someone’s homestead is archaic in German and we definitely want to avoid it here. *Burg* is the German word for *castle* in the sense of a fortress, palace, or fortified stately mansion that used to be occupied by a ruler. It is the best choice out of the alternatives, *Schloss* (*stately mansion*), *Festung* (*fortress*), *Palast* (*palace*), *Palais* (a smaller palace in French style). Since historically stately mansions were often fortified, we also quite frequently find the compound noun *Burgschloss* or *Schlossburg*, both meaning the same thing, a fortified stately mansion. The existence of these two synonymous compound nouns results in both *Burg* and *Schloss* being synonymous short forms of these compounds in many contexts, but since a *Schloss* need not necessarily be fortified, I thought it best to go with *Burg* as the translation for *castle*, as it applies to both castles and fortified stately mansions in German. Conceptually the image of a home as a castle invokes the protective aspects of a castle: a home is a place to retreat to from the world and its toils and worries. In the spirit of this conceptual image it is noteworthy to point out that English and German share a related metaphor: *to be the lord of the manor* (*der Herr im Haus(e) sein*). *Herr* is the polite form of address for men in modern German, analogous to English *Mr(.,)*, but historically it finds its roots in the formal address for noblemen which is where the connection with English *lord* lies. A *manor* is a stately mansion which, as I said, is also one of the readings of *Schloss* in German, which is in the same semantic field as *Burg* (*castle*) and sometimes used synonymously. Hence, if German speakers make this conceptual connection the English metaphor would be intelligible to them.

(21a) Truth is stranger than fiction.

(21b) *Die Wahrheit ist komischer als die Fiktion.*
 The truth is stranger than the fiction.
 ‘Truth is stranger than fiction.’

Strange in English and its counterpart *komisch* in German have two distinct senses: (1) uncanny, funny, and (2) unfamiliar, abnormal, weird. In English, *strange* can mean *funny* in the sense of odd or uncanny, but it cannot mean *funny* in the sense of provoking amusement, whereas in German *komisch* can be used in the sense of provoking amusement. We might take the metaphoric analogy to invoke either one of the two senses (uncanny and odd): truth, meaning reality, can be more abnormal than fictitious facts and events. The coincidentalness of this might strike us as funny. Imagine, for instance, a case where a surprise birthday party is being planned. The birthday boy or girl has grown suspicious of the secretive behaviour of their friends and relatives and have started to imagine all sorts of scenarios of what might be going on. When they are finally surprised with the party, they burst out into laughter because they are so much more relieved by what was truly going on as opposed to the scenarios they imagined. We might even take it to invoke both senses equally: take the previous example and add the assumption that the birthday boy or girl is laughing about how foolish it was of them to imagine highly unlikely scenarios rather than coming to the most obvious conclusion. They may find this fact funny and abnormal at the same time in a self-deprecating way. Both *truth* and *fiction* have the same meaning in English and German and the compositional meaning is the same due to the same syntax in both languages. The presence of a definite determiner in the two noun phrases *die Wahrheit* and *die Fiktion*, in this case, is not because the noun *Wahrheit* and *Fiktion* were count-nouns in German that required the presence of a definite determiner, but is added intentionally because its presence invokes a more general sense of the words. Compare how the definite determiner in the phrase *the love I feel* gives the noun *love* a more specific sense than in the phrase *love is all around*, in *the love I feel* we have a particular relationship in mind whereas in *love is all around* we are talking about all relationships. Similarly, the presence of the definite determiner in the German translation invokes a more specific mental scenario which should make the conceptual mapping implied by the metaphoric analogy more tangible for German speakers.

(22a) Beauty is skin-deep.

(22b) *Schönheit ist hauchdünn.*
 Beauty is very thin.
 ‘Beauty is skin-deep.’

Beauty has exactly the same meaning in English as it does in German. Notice that the German translation uses the adjective *hauchdünn*. Literally *hauchdünn* is composed of *hauch* and *dünn*. *Dünn* means thin and has exactly the same meaning as English *thin*. *Hauch* means a breath of air. Notice that I have glossed *hauchdünn* with the translation very thin. We can derive this as a compositional meaning for the German adjective: we usually think of air as nothing or the absence of something (e.g. the hole in the proverbial cheese). If you have ever prepared slices to be examined under a microscope you might have experienced that the thinner the slices are, the more translucent the material gets. Below a certain thickness the slice will be completely translucent, just like air. Hence, conceptually a slice that is as thin as a breath of air because of the translucency and the fragility of the slice. Skin layers have the same property. Each layer on its own is very thin and in some parts of the body exhibit the same translucency, e.g. the ears and the skin between the fingers and toes. We can therefore conclude that the German adjective *hauchdünn* and the English *skin-deep* are quite similar conceptually. The adjective *hauchdünn* can also be understood to evoke an image of thinness per se, rather than relating to

the skin. This sense of *hauchdünn* is relevant for the English proverb *beauty is skin-deep* which acknowledges that beauty is perceived on the surface. We can also take *beauty is skin deep* to evoke a sense of BEAUTY being perceived at the level of the skin, which is absent from the translation.

(23a) Custom is a second nature.

(23b) *Gewohnheit ist eine zweite Haut.*
 Custom is a second skin.
 ‘Custom is a second nature.’

Custom and *Gewohnheit* have the same meaning in English and German as *Gewohnheit* can morphologically be decomposed into the stem *Gewohn*, related to the adjective *gewohnt* (*being accustomed to*) which in turn is derived from the participle form of the verb *ge-wohn-t* (note the circumfix *ge- -t*, similar to Old English participle forms). The derivational suffix *-heit* is similar to English *-hood*, meaning the general state of being what the stem denotes. Hence we can compose the meaning of the German noun *Gewohnheit* conceptually as *the general state of being accustomed to something*, which is equivalent to the meaning of custom in the context of the English metaphor. I opted to translate *nature* with *Haut* (*skin*) because the phrase *zweite Haut* has a conventional fixed expression in German which has the idiomatic meaning that is the conceptual equivalent of the English idiomatic expression *second nature*. We can easily see the conceptual analogy: the essence of a person is what is in that person’s nature. This essence is as bound to that person as is that person’s skin. Also, in other German idiomatic expressions we find *Haut* associated with that which is characteristic for a person or as a metonymy of that person, as in *mit Haut und Haaren* (the wholeness of a person, literally *with skin and hair*). I think we can therefore conclude that the German translation of the metaphor comes as conceptually close as possible to the English original. There are other idiomatic expressions in German that relate to *custom* used in the sense of the English metaphor, e.g. *sich auf die faule Haut legen* (literally *to lie down on, to rest on one’s lazy skin*) even contains the word *Haut* (*skin*) in its idiomatic sense where it metaphorically stands for the person as a whole and the phrase *faule Haut* (*lazy skin*) alludes to that person’s nature, their character. Given these similarities across the idioms in German we might expect that if German speakers make these conceptual connections the English metaphor *custom is a second nature* could be intelligible to them.

(24a) Praise is not pudding.

(24b) *Lob ist kein Pudding.*
 Praise is no pudding.
 ‘Praise is not pudding.’

Praise & German *Lob* and *pudding* have similar meanings in English and German; however, in English, *pudding* has the sense of dessert in general, which it does not have in German. Note that the negation in German is nominal (whatever praise is, that thing is not pudding) while in English it is a verbal negation (the analogy between praise and pudding is being negated), but I would argue that these two interpretations equate to the same metaphoric conceptual meaning. Also, the use of the nominal negation *kein* is highly frequent and conventional in German in many idiomatic expressions such as *das ist kein Zuckerschlecken* (*that’s no cakewalk*, literally

that is no sucking-sugar) which should make this English metaphor seem more familiar to German speakers.

(25a) Promise is debt.

(25b) *Ein Versprechen ist eine Schuld.*
A promise is a debt.
'Promise is debt.'

This is another old proverb. *Promise & Versprechen* and *debt & Schuld* have exactly the same meanings in English and German. However, notice that in the German translation *Versprechen* occurs with the indefinite determiner *ein* (*a*) while in the English original *promise* occurs as a bare noun phrase, a noun phrase without a determiner. *Promise* in English can also be countable as in *I made her a promise*. *Versprechen* can only occur in a bare noun phrase when it is plural or, to put it differently, it can occur without a definite or indefinite determiner, but when it does it is not a mass noun (non-countable noun), but a countable noun in plural; the plural happens to be morphologically identical to the singular form. *Schuld*, on the other hand, can be used as a non-countable noun in German where it is singular and not plural, e.g. *heute wird über Schuld oder Unschuld entschieden* (in a court case, *today, the guilt or, literally, non-guilt will be decided*). While Old English *gylltas*, depending on context, could mean debt, guilt, or sin, modern English *debt* applies only to debt in the financial or moral sense while in modern German it has the same three senses as in Old English, i.e. it can refer to a moral obligation in a specifically religious context. In modern German, when *Schuld* is used in any context that is not explicitly religious, the financial sense would be the most prominent default sense. Therefore, *Schuld* is a more appropriate word choice in the translation than *Sünde* (*sin*). Note, however, that the noun *Schuld* in modern German can refer to both debt in the financial sense and guilt in the judicial sense, and both of these senses are relevant in the context of the metaphor. So by choosing the noun *Schuld*, German speakers should be more likely to infer the metaphoric meaning closest to the English original. Conceptually, it should be morally clear to German speakers that a promise can be understood as a kind of verbal contractual agreement that is only concluded upon its fulfilment and until this conclusion is reached the unresolved promise imparts a *Schuld* on the party bound by the promise.

German-specific metaphors

(26a) Dreams are froth.

(26b) *Träume sind Schäume.*
Dreams are froth.
'Dreams are froth.'

Most German speakers will probably identify this proverb as a quote of Sigmund Freud and indeed when we look at the diachronic corpus frequency of *Träume sind Schäume* in Figure (A.2) in Appendix A, there is a distinct spike around 1911 when Freud's first widely received book was published, but we also see that it was already in use before 1911 and its frequency continues to rise steadily to this day. Most German speakers today would be familiar with the metaphor, but few would be aware that it was also used by Sigmund Freud. The word *dream* means exactly the same thing in English as it does in German: the imaginary and

often imaginative mental sceneries experienced during REM sleep or the act of imagining such experiences as part of a longing, ambition, or aspiration, and it may refer to foolish wishful thinking as well. For the translation of *Schäume* we have two alternatives in English: *froth* and *foam*. I have decided to go with *froth* because this is the word used in translations of Freud's work in the English-speaking world, which is how some English speakers might have come across it. Also, in English *foam* can also be used for non-natural kinds of foam such as the foam used in construction work. *Froth*, on the other hand, is necessarily natural. So although we are interested in finding language-specific metaphors that are unfamiliar to monolinguals of the other language, this metaphor could have been familiar to them. As we will see in the experiments, English monolinguals were not familiar with it, probably because Freud's work is not read much at the beginning of the 21st century. Conceptually, English speakers might be able to draw a connection between dreams and froth. Dreams are often fleeting, for instance, the memory of a dream might be fresh in one's mind right after we wake up, but after that it often fades quickly. Similarly, if we take *dream* in the sense of a pipe-dream or wishful thinking, this sense also includes an element of temporal elusiveness: a person might be very enthusiastic about a dream one minute and might have given up on it in the next, or a person might give up on their dream, their aspiration, upon learning that it is a utopian impossibility. *Froth* also has this aspect of non-permanence: the image of froth caused by the surf pounding against a rock on the sea coast comes to mind, where the froth forms and disappears with the next wave. Or we may think of soap bubbles that do not last forever and notice how when we think of bubbles we also think of them as fragile. Hence, both the aspect of non-permanence and fragility should be conceptually apparent to English speakers, in which case they should be able to see that these are the properties that *dreams* and *froth* have in common.

(27a) Art is a science.

(27b) *Kunst ist eine Wissenschaft.*

Art is a science.

'Art is a science.'

German *Kunst* and English *art* share most of their sense: it can be a piece of art or artwork, a collection of such pieces or an exhibition thereof, it is the craft of making such pieces of art, the artistry. However, in German it can also refer to anything that is artificial as opposed to being a part of nature or created by nature, dating back to Latin *artifex*. Note, however, how this different sense also applies to pieces of art and could therefore also be incorporated into the metaphoric meaning. *Science* and *Wissenschaft* have the same meaning in English and German, at least for non-academics. Notice that morphologically, the German word *Wissenschaft* consists of the stem *Wissen* (*knowledge*) and the derivational suffix *-schaft* which is similar to English *craft*. In German, the suffix means the craft of making that which is denoted by the nominal stem, hence, *Wissenschaft* literally translates to *the making of knowledge/the craft that creates knowledge*. In German academia there is usually a finer distinction into *die Geisteswissenschaften* (*the humanities*, literally *the sciences of the mind*), *die Sozialwissenschaften* (*the social sciences*), and *die Naturwissenschaften* (*the natural sciences*) which mirror the same distinction in English academia, but note that some German and English academics might take the term *science* (*Wissenschaft*) to only apply to *the natural sciences* (*die Naturwissenschaften*). The fact that the humanities use the term *science* in the German academic term might mean that some speakers have a wider definition of *Wissenschaft* than

science in English, but notice how the compositional meaning of *Wissenschaft* as *the craft of making knowledge* builds on the meaning of *knowledge*, which I argued is the same in English and German. In this wide sense, German non-academics, who are not invested in the debate whether the humanities are fundamentally different from the natural sciences, might be inclined to say that the humanities do contribute to our body of knowledge and thus are *sciences* in this broad sense. This wide, compositional sense of *Wissenschaft* is often exploited in a play on words and used by universities and university towns in advertisement with slogans such as *die Stadt, die Wissen schaft* (*the town that creates knowledge*), which shows that the average German speaker is well-aware of this broad interpretation. Also notice how in both English and German *science* (*Wissenschaft*) can be used in idiomatic expressions, such as *it's not rocket science* or *mach keine Wissenschaft draus* (literally *don't make a science out of it*), where it conveys in both languages that the addressee should not act as if something was difficult, i.e. requiring the sort of learning associated with academia. These idiomatic expressions suggest that speakers of both languages can use *science* (*Wissenschaft*) in a broad sense. We might therefore conclude that the lexical and metaphoric meanings are cross-linguistically comparable and quite similar if not the same.

(28a) Beauty is a ticket.

(28b) *Schönheit ist eine Eintrittskarte.*

Beauty is a ticket.

‘Beauty is a ticket.’

I already noted in the discussion of the cross-linguistically shared metaphors that *beauty* has exactly the same meaning and connotations in English and German. The German noun *Eintrittskarte* is a compound noun consisting of *Eintritt* (*entering*) and *Karte* (*card*), which gives as the compositional conceptual meaning of *a card that permits one to enter*. It is important to note that there are different types of compound nouns with *Karte* as one of the nouns for different kinds of tickets: *Eintrittskarte* is, for instance, used to refer to *theatre tickets*, *baller tickets*, or *tickets for the cinema*, while tickets for any kind of transportation (bus, train, ship, ferry, plane) would be *Fahrkarte* (*transportation card*, literally *driving card*). Also, recently *ticket* has been borrowed into German and has retained its full English meaning and is therefore indiscriminately used for all types of tickets. In the conventional German expressions *Eintrittskarte zum Ruhm* (*ticket to fame*) or *Eintrittskarte zu den oberen Zehntausend* (*ticket to the upper ten-thousand meaning the rich*) we also find the noun *Eintrittskarte* rather than *Fahrkarte* being used. Since there is no conventional compound nouns with the specific meaning of *Eintrittskarte* in English I opted for the more comprehensive term *ticket*. And because the noun *ticket* is broader and more inclusive, the mental image that it invokes should be less restricted which should mean that it would be easier for English speakers to find potential analogies between the social status of beauty and the cultural conventions that permit someone acceptance into certain social circles.

(29a) Election day is pay day.

(29b) *Wahltag ist Zahltag.*

Election day is pay day.

‘Election day is pay day.’

This is a relatively new saying. We can see this from the diachronic corpus frequency shown in Figure (A.2) in Appendix A. The reason for its recency finds its roots in history. The metaphor's diachronic trajectory corresponds well with the rise and fall and rise of democracy in 20th century Germany: we find its first occurrences during the time of the parliamentary monarchy and representative democracy of the Weimar Republic, 1919-1930, then again in the liberal democracy, 1930-1933, where chancellor Hindenburg was a *de facto* dictator ruling depression-era Germany with emergency legislative and executive powers. Simultaneously, Nazi dictatorship had already started to take over and interfered with elections until 1933. We find more occurrences of the metaphor after World War II, during the time of the German republic of West Germany, 1949-1989/90 and in communist East Germany, and finally in the re-united republic, East and West, 1989/90 to present. Throughout its history, we see that the diachronic trajectory of the metaphor continues to rise. Only since re-unification has started to decline. *Election day & Wahltag* and *pay day & Zahltag* have exactly the same meanings in English and German and our modern-day English and German experimental participants would have grown up with similar cultural ideas about free elections, the economy, and the welfare state.

(30a) Sport(s) is murder.

(30b) *Sport ist Mord.*
Sport(s) is murder.
'Sport(s) is murder.'

This metaphoric proverb allegedly goes back to a statement made by Winston Churchill during an interview. However, after I did some research, this back-story turns out to be a hoax created by a reporter looking to advance their career at the time. Nevertheless, this story keeps being perpetuated throughout the German-speaking world, lending false credibility to the proverb, but as we can see from the diachronic corpus frequency in Figure (A.2) in Appendix A, the proverb continues to increase in popularity. The meaning of *murder* is exactly the same in English and German. The meaning of *Sport* in German is the same as in English, apart from the fact that it cannot be used to refer to a person as *an old sport*, but this sense seems less relevant in the context of the metaphor. Note that I have put the s in sport(s) in parenthesis to indicate that corpus frequencies were checked for both British English and American English spellings. The monolinguals in the experiments are predominantly British English speakers and saw the appropriate spelling. The idiomatic meaning of this metaphor is that physical exercise is straining and exhausting (which should also be conceptually plausible to English speakers) but specifically for those who are non-athletic.

(31a) The devil is a squirrel.

(31b) *Der Teufel ist ein Eichhörnchen.*
The devil is a squirrel.
'The devil is a squirrel.'

This is a very old proverb that goes at least as far back as at the Middle Ages. The implied metaphoric analogy between the devil and a squirrel is twofold: (1) in terms of behaviour: misfortune strikes suddenly and unexpected and the devil is the embodiment of that misfortune. Squirrels jump around erratically in a similarly surprising way. (2) visually: the European squirrel is crimson-coloured. Note that the grey squirrel that is found alongside the crimson

squirrel in Europe now originates from the Americas, hence Medieval Germans would have only known crimson-coloured squirrels and this explains how a squirrel is also visually similar to the devil. With the rise of atheism, this metaphor has begun to decline, but keep in mind that Medieval Germans might have taken it as an omen of impending misfortune, much like sightings of snakes were interpreted as omens because the serpent is the Christian symbol of sin while sightings of black cats might be good or bad depending on the direction in which the cat is crossing one's path. The words *devil* and *squirrel* have exactly the same meanings in English and German and speakers of both languages should be familiar with the role of the devil in the Christian faith and its cultural heritage.

(32a) Following is suffering.

(32b) *Mitgehangen ist Mitgefangen.*
 Hang-alongside is caught-alongside.
 'Following is suffering.'

This is a very old metaphoric proverb and is still very much in use. However, over time the phrasing has changed. Originally the proverb was *mitgegangen, mitgefangen, (und) mitgehangen* and talks about the fate of those who are found in the company of criminals. Those who *go along with* (*mit-gegangen*, literally the past tense of *go-along*, the prefix *mit-* means *with* or *to follow*) criminals and are caught with them (*mit-gefangen*, literally the past tense of *caught-with*) also end up being hanged for the same crimes (*mit-gehangen*, literally the past tense of *hang-with*) because they are assumed to be co-conspirators. In the modern phrasing the *mitgegangen* has disappeared and *mitgefangen* and *mitgehangen* have changed places, which should imply a conceptually different meaning than the original phrasing; however, according to the Duden, the official German dictionary, the proverb has retained its original meaning. The meaning of *mitgehangen* has also widened: it no longer applies to hanging, but any sort of punishment. Essentially the meaning of the proverb has been widened to the point where it means *those found with the accused are likely to be suspected of the same accusations*. I therefore decided against a very literal translation here and opted for one that is more in the spirit of the conceptual imagery, for two reasons: (1) a lot of the metaphoric meaning is expressed by use of morphology in the German original and this cannot faithfully be duplicated in English with phrasal verbs such as *hang-with* and *caught-with*, and (2) the reversed order of *mitgehangen* and *mitgefangen* is ignored by German speakers who, because of the idiomatic meaning, know that conceptually it needs to be the other way around and so this word order would have just confused speakers of English conceptually.

(33a) Hope is patience.

(33b) *Hoffnung ist Geduld.*
 Hope is patience.
 'Hope is patience.'

Hope & Hoffnung and *patience & Geduld* have exactly the same meanings in English and German. Conceptually, English speakers should be able to see that (1) *hope* requires *patience*, but (2) *hope* also implies a holding out, a patient waiting, and seeing these two conceptual analogies could make the German metaphor intelligible to speakers of English.

Table 6.2: Relative Levenshtein distance of compositional semantics. The compositional Levenshtein distance of a metaphor is $\%Lev = (Lev_1 + Lev_2)/2$, where Lev_1 is the distance of the metaphor source and Lev_2 of the metaphor target, given here per metaphor type: English-specific (E), German-specific (G), cross-linguistically shared (S), and novel (N).

Lev	%E	%G	%S	%N
0	33.3	25	30.8	50
0.2	50	25	38.5	33.3
0.4	0	33.3	23.0	8.4
0.6	16.7	8.4	7.7	8.3
0.8	0	8.3	0	0
1	0	0	0	0

For the language-specific metaphors, we can argue that their lexical and compositional semantics should be comparable between English and German and that the only semantic difference should be that native speakers will have idiomatic meanings associated with them whereas non-native speakers will not: (1) the idiomatic meanings of English-specific metaphors will be familiar to English monolinguals but not German monolinguals, and (2) the idiomatic meanings of German-specific metaphors will be known to German monolinguals but not to English monolinguals.

Novel metaphors

As for the novel metaphors, since there are no idiomatic meanings which speakers of either language can access we would have to assume that to say that the novel metaphors cross-linguistically have comparable meanings is to say that, for all intents and purposes, they are comparable in terms of lexical and compositional semantics. As we can see from the relative Levenshtein distances in Table 6.2 half of the novel metaphors should have the same compositional meaning in English and German with another 1/3 having almost the same compositional meaning (a Levenshtein distance of less than .2) and this distribution of relative Levenshtein distances is comparable to that of the other metaphor types.

In order to quantify the semantic Levenshtein distance for a metaphor we can therefore neglect the copula and simply assume that the distance between the English and German translation of the metaphor source is added to the translation distance of the metaphor target and then we normalise again by dividing by the number of concepts, two. If (1) we could agree on how much metaphor source and target as well as *ad hoc* concepts and emergent properties contributed to the overall metaphoric analogy and if (2) we could also assume that experimental participants would also infer the same metaphoric meaning (or a reasonable approximation of mental states) and in a similar or the same way (or, again, in approximation thereof), then it would be advisable to weigh the contributions accordingly in the formula. However, I do not think we understand enough about the magnitude of their contributions, let alone ways to measure them yet and so I do not think a weighted calculation is warranted yet. When we examine the distributions of compositional Levenshtein distances per metaphor type in Table 6.2, we find that although the distributions are a bit more uneven while the lexical Levenshtein distances had a clear trend to be smaller rather than greater, (1) the trend of distances being

smaller rather than greater is still generally true and (2) the Levenshtein distributions are comparable across metaphor types.

The cross-linguistically shared metaphors have well-established wordings in the two languages, English and German, and we want to use these wordings because these are the ones that are conventionally accepted and thus most familiar to speakers and also the ones that they will expect. Indeed, any attempt to present native speakers with translations that deviate from these conventional wordings might confuse them and lead them to find them less plausible. In order to assess the cross-linguistic intelligibility of L2 metaphors we are comparing the plausibility ratings native speakers give in the judgement task to their ratings of the L1 and shared metaphors. Hence, if we caused participants to find the cross-linguistically shared metaphors less plausible because we changed their wording, we would make them a less suitable control condition against which to evaluate the L2 metaphor ratings. In my discussion of the cross-linguistically shared metaphor I have therefore pointed out these conventionally accepted wordings, but I have also tried to show how they are conceptually comparable between English and German.

For language-specific metaphors native speakers have similarly high expectations of conventional wordings of the L1 metaphors, the metaphors from their native language, and so we want to use these. Since the language-specific metaphors and their idiomatic meanings will be unfamiliar to L2 speakers, native speakers of the other language, we want to ensure that the translation of the language-specific metaphors are as close to the original as possible. I have argued for this in two ways: (1) I have looked at the comparability of the lexical meanings and argued that they are cross-linguistically comparable, and (2) I have tried to show how, even though the L2 speakers would not know the idiomatic meanings of the metaphors, they might still infer a metaphoric interpretation for them that comes close to it by (a) considering other idiomatic expressions in their L1 that conceptually relate to the metaphoric meaning in the L2 and (b) by considering the senses of the metaphor source and target concepts that they would have from their L1 and that would be relevant in the context of the metaphors.

It is especially important to point out these conceptual similarities because they are of the sort that Conceptual Metaphor Theory argues are necessary for speakers to infer the meaning of complex linguistic metaphors from basic conceptual metaphors. I have argue that speakers of both languages should have conceptual analogies for the cross-linguistically shared and language-specific metaphors to make them cross-linguistically intelligible.

Summary: quantifying the cross-linguistic comparability of the material

All cross-linguistic differences in linguistic form, familiarity, and meaning that we have looked at in this chapter are summarised in Table (6.3). In general we can see from this table that the linguistic form of the metaphors and familiarity of that form is no more than 12% dissimilar between the English and German translation of the experimental material. We also see that the variation in terms of the linguistic form and its familiarity is, on average, equally distributed across the four different metaphor types. As for the linguistic meaning of the metaphors we see that they are no more than 19% dissimilar between English and German and, as with their form and familiarity, the variation in meaning dissimilarities is on average equally distributed

across the metaphor types.

Table 6.3: Cumulative differences between the English and German version of the experimental material in terms of their linguistic form and meaning.

Form	E	G	S	N
Synchronic metaphor freq.	6.0%	6.0%	4.5%	NA
Diachronic m. freq. change	4.8%	1.8%	5.5%	NA
Synchronic word freq.	0.5%	2.2%	1.0%	1.5%
Metaphor length # char.	20.9%	21.5%	15.8%	17.7%
# syll.	22.3%	27.5%	22.0%	12.5%
# morph.	8.9%	10.2%	3.2%	0%
# words	20.8%	12.5%	7.1%	0%
Cumulative formal difference	12.0%	11.7%	8.4%	6.3%
Meaning	E	G	S	N
Lexical semantics	13.3%	22.5%	14.5%	14.3%
Contextual relevance of diff.	14.0%	11.0%	14.0%	0%
Comp. semantics	13.1%	23.5%	15.2%	10.7%
Cumulative sem. difference	13.5%	19.0%	14.6%	8.3%

The columns in Table (6.3) show the type of difference in the first column, the relative difference per metaphor type—English (E), German (G), shared (S), and novel (N)—in columns three through six. The row *cumulative formal difference* is the average of all rows above it, except those marked NA. The row *cumulative semantic difference* is the average of all rows above it concerning semantics. Table (6.3) thus allows us to look at two things: (1) it allows us to gauge *how similar* the English and German version of the experimental material is in terms of its linguistic form, familiarity of that form, and its linguistic meaning, and (2) it allows us to check whether the dissimilarities that do exist between the English and German version of the material are evenly distributed across the four metaphor types.

As for (1), we see that the translations of the material are much more similar than dissimilar: the translations, on average, are 88.3% similar, i.e. 90.4% similar in terms of their linguistic form and familiarity of that form and 86.2% similar in terms of their lexical and compositional semantics and contextual relevance of differing senses. As for (2), we find that overall the formal and semantic translation differences are evenly distributed across the metaphor types: the only difference with a variation of $p < .05$ is metaphor length measured as the number of syllables and the second lowest p -value is .065 for differences in metaphor length measured as the number of morphemes. Both the differences in the number of syllables and morphemes, can easily be explained by pointing out that German is richer in morphology (see the ratio of affixes to stems in Panel D in Figure 5.7) and those orthographic letters at the end of words that would be silent in English (e.g. the *e* at the end of the noun *love*) are pronounced in German (e.g. *Liebe*, love) and since these sounds are often vowels they constitute additional syllables, thereby increasing the number of syllables in the German translation compared to the English version of the material. The differences in the number of syllables and morphemes is thus systematic with respect to differences in phonology and morphology. For the other differences, notice that the p -values are all above .24 and several close to 1. We can therefore conclude that, for all intents and purposes, the variation in relative differences between metaphor types is evenly distributed.

What follows next is a detailed breakdown of how all of the values given in Table (6.3) were calculated.

First half of the table: linguistic form

First row: The synchronic metaphor frequency difference ΔF is the difference of the cumulative metaphor probabilities given in brackets in Table (5.2). The cumulative frequency difference for the English metaphors ΔF_E ($F_{AE} + F_{BE}/n$, where $n = 2$) and German metaphors ΔF_G is calculated in the same way:

$$\Delta F_E = \Delta F_G = |F_E - F_G| = \left| \left(\frac{F_{AE} + F_{BE}}{n} \right) - F_G \right| = \left| \left(\frac{.92 + .24}{2} \right) - .64 \right| = .06 = 6\% \quad (6.2)$$

Note that the cumulative differences for the American English F_{AE} and British English corpora F_{BE} are averaged before comparing them to the cumulative difference in the German corpora F_G . Similarly for ΔF_S :

$$\Delta F_S = |F_E - F_G| = \left| \left(\frac{F_{AE} + F_{BE}}{n} \right) - F_G \right| = \left| \left(\frac{.98 + .85}{2} \right) - .96 \right| = .045 = 4.5\% \quad (6.3)$$

Since the novel metaphors do not occur in any of the corpora, it makes no sense to calculate a relative difference in cumulative probability. The cell is thus marked NA (not applicable). As noted earlier, the cumulative probability can be understood as the likelihood of a metaphor to be familiar. Therefore, the difference in cumulative probability can be interpreted as the relative difference of that likelihood.

Second row: Diachronic metaphor frequency change is the relative change in absolute cumulative probability between the present-day probability and the probability 30 years ago. **Third row:** Synchronic word frequency. See Table (A.2) for the p -values used in the calculation. Note that some words occur not just in one but multiple metaphor types. In Equations (6.4a) to (6.4d), the average cumulative difference ΔF is the difference between the cumulative probability of a word in English and the word used in the German translation for it, $|F(w_E) - F(w_G)|$, averaged across all translation pairs (English–German) for all metaphors, but for each metaphor type individually. The cumulative probability of the English word $F(w_E)$ is the cumulative probability for that word derived from its frequency in the English corpora. The cumulative probability of the German word $F(w_G)$ is its cumulative probability based on its frequency in the German corpora. The number of those words that occur in a metaphor type is $n_E(w)$ for English metaphors, $n_G(w)$ for German, $n_S(w)$ for shared, and $n_N(w)$ for novel metaphors. Since the cumulative probability can be interpreted as the likelihood of a metaphor to be familiar, the difference in cumulative probability can be understood as the relative difference of that

likelihood.

$$\Delta F_E = \frac{\sum_{i=1}^{n_E(w)} |F(w_E) - F(w_G)|}{n_E(w)} = \frac{|0.498 - 0.498| + \dots + |0.537 - 0.501|}{14} \approx 0.5\% \quad (6.4a)$$

$$\Delta F_G = \frac{\sum_{i=1}^{n_G(w)} |F(w_E) - F(w_G)|}{n_G(w)} = \frac{|0.500 - 0.499| + \dots + |0.741 - 0.641|}{26} \approx 2.2\% \quad (6.4b)$$

$$\Delta F_S = \frac{\sum_{i=1}^{n_S(w)} |F(w_E) - F(w_G)|}{n_S(w)} = \frac{|0.500 - 0.499| + \dots + |0.497 - 0.499|}{25} \approx 1.0\% \quad (6.4c)$$

$$\Delta F_N = \frac{\sum_{i=1}^{n_N(w)} |F(w_E) - F(w_G)|}{n_N(w)} = \frac{|0.500 - 0.499| + \dots + |0.570 - 0.499|}{17} \approx 1.5\% \quad (6.4d)$$

Fourth row: Metaphor length measured as the number of orthographic characters. See Panel A in Figure (5.7). Equations (6.5) to (6.8) are calculated for each metaphor type individually. The relative difference in the number of orthographic characters $\Delta L(c)$ is the difference $|n_E(c) - n_G(c)|$ in the number of characters in the English translation of each metaphor $n_E(c)$ and the German version of that same metaphor $n_G(c)$ relative to, i.e. divided by the number of characters of the metaphor with more characters in the English–German pair, $\max(n_E(c), n_G(c))$. This difference per metaphor is then summed for each metaphor type $\sum_{i=1}^j \Delta L$ and averaged to the number j of metaphors of that type.

$$\begin{aligned} \Delta L_E(c) &= \left(\sum_{i=1}^{j_E} \left(\frac{|n_E(c) - n_G(c)|}{\max(n_E(c), n_G(c))} \right) \right) / j_E \\ &= \left(\left(\frac{|11 - 14|}{\max(11, 14)} \right) + \dots + \left(\frac{|13 - 27|}{\max(13, 27)} \right) \right) / 12 \\ &= \left(\left(\frac{3}{14} \right) + \dots + \left(\frac{14}{27} \right) \right) / 12 \approx .209 = 20.9\% \end{aligned} \quad (6.5)$$

$$\begin{aligned} \Delta L_G(c) &= \left(\sum_{i=1}^{j_G} \left(\frac{|n_E(c) - n_G(c)|}{\max(n_E(c), n_G(c))} \right) \right) / j_G \\ &= \left(\left(\frac{|14 - 17|}{\max(14, 17)} \right) + \dots + \left(\frac{|14 - 17|}{\max(14, 17)} \right) \right) / 12 \\ &= \left(\left(\frac{3}{17} \right) + \dots + \left(\frac{3}{17} \right) \right) / 12 \approx .215 = 21.5\% \end{aligned} \quad (6.6)$$

$$\begin{aligned} \Delta L_S(c) &= \left(\sum_{i=1}^{j_S} \left(\frac{|n_E(c) - n_G(c)|}{\max(n_E(c), n_G(c))} \right) \right) / j_S \\ &= \left(\left(\frac{|14 - 20|}{\max(14, 20)} \right) + \dots + \left(\frac{|15 - 15|}{\max(15, 15)} \right) \right) / 14 \\ &= \left(\left(\frac{6}{20} \right) + \dots + \left(\frac{0}{15} \right) \right) / 14 \approx .158 = 15.8\% \end{aligned} \quad (6.7)$$

$$\begin{aligned}
\Delta L_N(c) &= \left(\sum_{i=1}^{j_N} \left(\frac{|n_E(c) - n_G(c)|}{\max(n_E(c), n_G(c))} \right) \right) / j_N \\
&= \left(\left(\frac{|13 - 16|}{\max(13, 16)} \right) + \dots + \left(\frac{|14 - 18|}{\max(14, 18)} \right) \right) / 12 \\
&= \left(\left(\frac{3}{16} \right) + \dots + \left(\frac{4}{18} \right) \right) / 12 \approx .177 = 17.7\%
\end{aligned} \tag{6.8}$$

Fifth row: Metaphor length measured as the number of syllables. See Panel B in Figure (5.7). In analogy to Equations (6.5) through (6.8) we define metaphor length measured as the number of syllables $n(\sigma)$ as follows and calculate it for each metaphor type individually.

$$\Delta L(\sigma) = \left(\sum_{i=1}^j \left(\frac{|n(\sigma) - n(\sigma)|}{\max(n(\sigma), n(\sigma))} \right) \right) / j \tag{6.9}$$

Sixth row: Metaphor length measured as the number of morphemes. See Panel D in Figure (5.7). Note that the graph in Panel D distinguishes between stem, i.e. root morphemes, and affixes, while the calculation treats them equally. We define metaphor length measured as the number of morphemes $n(\mu)$ as follows in analogy to Equations (6.5) through (6.8) and calculate it for each metaphor type individually.

$$\Delta L(\mu) = \left(\sum_{i=1}^j \left(\frac{|n(\mu) - n(\mu)|}{\max(n(\mu), n(\mu))} \right) \right) / j \tag{6.10}$$

Seventh row: Metaphor length measured as the number of orthographic words per metaphor. See Panel C in Figure (5.7). Note that the graph in Panel C distinguishes between differences in the number of words caused by (1) mismatches in the mass-noun/count-noun distinction (e.g. *bliss* is used as a mass-noun in the metaphor *ignorance is bliss*, but in the German version it is a count-noun, *Unwissenheit ist ein Segen*, lit. *ignorance is a blessing*), (2) mismatches in the presence or absence of a determiner that yield a poetic or generic reading (e.g. *the love we share* versus *love is all around*), and (3) purely orthographic differences (e.g. compound nouns such as *election day* and *pay day* are spelled as two orthographic words in English, but as one orthographic word in German, *Wahltag* and *Zahltag*). In the calculation all of these differences are quantified equally. In analogy to Equations (6.5) to (6.8) metaphor length measured as the number of words $n(w)$ is defined as follows and we have to calculate it for each metaphor type separately.

$$\Delta L(w) = \left(\sum_{i=1}^j \left(\frac{|n(w) - n(w)|}{\max(n(w), n(w))} \right) \right) / j \tag{6.11}$$

Second half of the table: linguistic meaning

First row: Lexical semantics. See Table (A.5) for all values and Table (6.1) for the variation within each metaphor type. The relative semantic Levenshtein distance, lev , of a concept can be understood as the percentage of semantic overlap. It is the number of senses that would have to be added, taken away, or changed to transform the English concept C_E in each English–German translation pair into its German counterpart C_G , normalised relative to the total number of senses in the concept with more senses in each pair. This distance is therefore the proportion of sense differences between the English and German metaphor concepts. For each metaphor

source and target concept it is calculated using Equation (6.1) and normalising afterwards. The distances lev_x for each English–German translation pair x is summed for each metaphor type individually and averaged by dividing by the number n of translation pairs for that metaphor type. Note that some concepts occur in multiple metaphor types.

$$\text{lev}_E = \frac{\sum_{x=1}^{n_E} \text{lev}_x(C_E, C_G)}{n_E} = \frac{0 + 0 + \dots + .4}{14} \approx .133 = 13.3\% \quad (6.12a)$$

$$\text{lev}_G = \frac{\sum_{x=1}^{n_G} \text{lev}_x(C_E, C_G)}{n_G} = \frac{.25 + 0 + \dots + .4}{26} \approx .225 = 22.5\% \quad (6.12b)$$

$$\text{lev}_S = \frac{\sum_{x=1}^{n_S} \text{lev}_x(C_E, C_G)}{n_S} = \frac{.25 + 0 + \dots + 0}{25} \approx .145 = 14.5\% \quad (6.12c)$$

$$\text{lev}_N = \frac{\sum_{x=1}^{n_N} \text{lev}_x(C_E, C_G)}{n_N} = \frac{0 + 0 + \dots + .4}{17} \approx .143 = 14.3\% \quad (6.12d)$$

Second row: Contextual relevance of the differing senses given in the previous row (lexical semantics). See Table (A.5) for a complete list. The values given in Table (6.3) are the ones in brackets in the headers in Table (6.1).

Third row: Compositional semantics. See Table (A.5) for all values and Table (6.2) for the variation within each metaphor type. Recall, I suggested to calculate the compositional semantic overlap of a metaphor as $\text{lev} = (\text{lev}_1 + \text{lev}_2)/2$, where lev_1 is the distance of the metaphor source in an English–German translation pair and lev_2 is the distance of the metaphor target respectively. To expand Equations (6.12a) through (6.12d) to compositional semantics, we replace $\text{lev}_x(C_E, C_G)$ with $(\text{lev}_1 + \text{lev}_2)/2$.

$$\text{lev}_E = \frac{\sum_{x=1}^{n_E} ((\text{lev}_1 + \text{lev}_2)/2)}{n_E} = \frac{0 + .083 + \dots + .417}{12} \approx .131 = 13.1\% \quad (6.13a)$$

$$\text{lev}_G = \frac{\sum_{x=1}^{n_G} ((\text{lev}_1 + \text{lev}_2)/2)}{n_G} = \frac{.333 + .125 + \dots + 0}{12} \approx .235 = 23.5\% \quad (6.13b)$$

$$\text{lev}_S = \frac{\sum_{x=1}^{n_S} ((\text{lev}_1 + \text{lev}_2)/2)}{n_S} = \frac{.25 + .2 + \dots + .125}{14} \approx .152 = 15.2\% \quad (6.13c)$$

$$\text{lev}_N = \frac{\sum_{x=1}^{n_N} ((\text{lev}_1 + \text{lev}_2)/2)}{n_N} = \frac{.333 + .417 + \dots + 0}{12} \approx .107 = 10.7\% \quad (6.13d)$$

In the previous chapter, I discussed the two types of significance (statistical significance and substantive significance) as well as ways to determine the probability of Type I and Type II errors in our statistics, Bayesian inference and two metrics of statistical reliability (the power of prior probabilities and the power of posterior probabilities). In the course of this discussion, I mentioned that many conventionally established recommendations for statistical quality standards exist: Fisher (1925) recommends that the chance of a test to yield Type I errors should be no more than 5%; Cohen (1988) assumes that, in most experimental settings, researchers would view Type I errors to be four times more serious than Type II errors. Adopting Fisher’s threshold for statistical significance, Cohen recommends that if the chance of Type I errors should be less than 5%, then the chance of Type II errors should be no more than 20%. The propensity of any test statistic to declare an effect statistically significant increases with sample size. However, despite the known risk of oversampling, to my knowledge, no recommendation as to a stopping point exists. We saw that because of this, statistical significance must always be interpreted in light of substantive significance, the actual size of the effect. Cohen (1992) provides a scale for interpreting effect sizes.

Table 6.4: Cross-linguistic variation in the experimental material. Shown are the number of samples N , the effect size (the difference in cumulative probabilities between English and German), the p -value of statistical significance (** for $p \leq .01$, † for $p \leq .1$), the probability of Type I errors α , the probability of Type II errors β , and statistical power $P = 1 - \beta$ ($\diamond = P < .8$). IDs: (1) Metaphor frequency, (2) diachronic frequency change, (3) word frequency, metaphor length measured as (4) the number of characters, (5) syllables, (6) morphemes, and (7) words, (8) lexical sense differences, (9) contextual relevance of sense differences, and (10) compositional-semantic differences. The last row in each table provides general estimates of central tendency, averaged across the individual studies and weighted according to their sample size.

English-specific metaphors					
ID	N	Effect size	$p = \alpha$	β	P
1	12	.060	.792	.073	.927
2	12	.048	.834	.068	.932
3	12	.005	.983	.052	.948
4	12	.209	.260	.170	.830
5	12	.223	.208	.183	.817
6	12	.089	.690	.088	.912
7	12	.208	.263	.169	.831
8	12	.133	.535	.113	.887
9	12	.140	.510	.118	.882
10	12	.131	.542	.112	.888
	120	.125	.562	.115	.885

German-specific metaphors					
ID	N	Effect size	$p = \alpha$	β	P
1	12	.060	.792	.073	.927
2	12	.018	.938	.056	.944
3	12	.022	.924	.058	.942
4	12	.215	.237	.175	.825
5	12	.275	.009 **	.236	.764 \diamond
6	12	.102	.645	.095	.905
7	12	.125	.564	.108	.892
8	12	.225	.086 †	.214	.786 \diamond
9	12	.110	.617	.099	.901
10	12	.235	.162	.194	.806
	120	.139	.497	.131	.869

Table 6.5: Cross-linguistic variation in the experimental material. Shown are the number of samples N , the effect size (the difference in cumulative probabilities between English and German), the p-value of statistical significance, the probability of Type I errors α , the probability of Type II errors β , and statistical power $P = 1 - \beta$. IDs: (1) Metaphor frequency, (2) diachronic frequency change, (3) word frequency, metaphor length measured as (4) the number of characters, (5) syllables, (6) morphemes, and (7) words, (8) lexical sense differences, (9) contextual relevance of sense differences, and (10) compositional-semantic differences. The last row in each table provides general estimates of central tendency, averaged across the individual studies and weighted according to their sample size.

Cross-linguistically shared metaphors					
ID	N	Effect size	$p = \alpha$	β	P
1	14	.045	.831	.069	.931
2	14	.055	.794	.074	.926
3	14	.010	.963	.054	.946
4	14	.158	.401	.140	.860
5	14	.220	.156	.198	.802
6	14	.032	.880	.063	.937
7	14	.071	.734	.082	.918
8	14	.145	.452	.130	.870
9	14	.140	.471	.126	.874
10	14	.152	.425	.135	.865
	140	.103	.611	.107	.893

Novel metaphors					
ID	N	Effect size	$p = \alpha$	β	P
1	12	NA	NA	NA	NA
2	12	NA	NA	NA	NA
3	12	.015	.948	.055	.945
4	12	.177	.377	.144	.856
5	12	.125	.564	.108	.892
6	12	0	1	.050	.950
7	12	0	1	.050	.950
8	12	.143	.499	.120	.880
9	12	0	1	.050	.950
10	12	.107	.627	.097	.903
	120	.071	.752	.084	.916

Table 6.6: Cross-linguistic variation in the experimental material, sorted from least to largest variation. Shown are the number of samples N , the effect size (the difference in cumulative probabilities between English and German), the p-value of statistical significance, the probability of Type I errors α , the probability of Type II errors β , and statistical power $P = 1 - \beta$. IDs: (1) Metaphor frequency, (2) diachronic frequency change, (3) word frequency, metaphor length measured as (4) the number of characters, (5) syllables, (6) morphemes, and (7) words, (8) lexical sense differences, (9) contextual relevance of sense differences, and (10) compositional-semantic differences. The last row in each table provides general estimates of central tendency, averaged across the individual studies and weighted according to their sample size.

	N	Effect size	$p = \alpha$	β	P
(3) Word freq	50	.013	.955	.055	.945
(2) Diachronic	50	.041	.852	.066	.934
(1) Meta freq	50	.054	.806	.072	.928
(6) Morphemes	50	.055	.807	.074	.926
(9) Relevance	50	.099	.642	.099	.901
(7) Words	50	.100	.644	.101	.899
(10) Comp sem	50	.156	.438	.135	.865
(8) Lex sem	50	.161	.395	.144	.856
(4) Characters	50	.188	.322	.157	.843
(5) Syllables	50	.211	.231	.182	.818
	500	.108	.609	.108	.892

Table 6.7: Cross-linguistic variation in the experimental material, averaged per metaphor type and weighted by sample size. Shown are the number of samples N , the effect size (the difference in cumulative probabilities between English and German), the p-value of statistical significance, the probability of Type I errors α , the probability of Type II errors β , and statistical power $P = 1 - \beta$. IDs: (1) Metaphor frequency, (2) diachronic frequency change, (3) word frequency, metaphor length measured as (4) the number of characters, (5) syllables, (6) morphemes, and (7) words, (8) lexical sense differences, (9) contextual relevance of sense differences, and (10) compositional-semantic differences. The last row in each table provides general estimates of central tendency, averaged across the individual studies and weighted according to their sample size.

Metaphors' linguistic form					
	N	Effect size	$p = \alpha$	β	P
E	84	.120	.576	.115	.885
G	84	.117	.587	.114	.886
S	98	.084	.680	.097	.903
N	84	.063	.778	.081	.919
	350	.096	.656	.102	.898

Metaphors' linguistic meaning					
	N	Effect size	$p = \alpha$	β	P
E	36	.135	.529	.114	.886
G	36	.190	.288	.169	.831
S	42	.146	.449	.130	.870
N	36	.083	.709	.089	.911
	150	.139	.492	.126	.874

The statistical power of a test is its rate of true positives: if there is positive evidence for an effect, then a reliable test would always identify the evidence as such. Following Cohen's recommendation, if the chance of Type II errors should be no more than 20%, then a test should give true positives at least 80% of the time. This metric of statistical power is well-established (albeit rarely used). In the previous chapter, we made a slight detour to Bayesian statistics and I showed that this power metric is the statistical reliability of *prior* probabilities, i.e. the experimental effect size estimates, and I introduced a method to determine the reliability of *posterior* probabilities found through Bayesian update of effect size estimates prompted by the measurement inaccuracy of the test. This proved to be a powerful set of tools in the corpus analysis and I suggest we subject the quantification of the material to it as well. There is another reason for doing this, despite meeting quality standards: if our rationale for approving the material "ready for experiments" because of some arbitrary quality threshold, then for the material that passed this bar, it would be as if we had expunged all record of whatever deficits remained and this could seriously hinder our ability to interpret the results.

Recall that in order to test the effect that non-conceptual aspects of language-specific idiomatic meanings may have on metaphor comprehension and speakers' consideration of meaning salience, we aim to set up a situation where, cross-linguistically, both English and German experimental participants, unbeknown to them, will be likely to make use of similar conceptual knowledge during comprehension. Already in Chapters 2 and 3 I cautioned that this can only be approximately achieved, mainly for two reasons: the language-specific metaphors we use must be translated and this translation could potentially be a source that introduces conceptual cross-linguistic differences into the experimental material. We intend to use metaphors that occur in both English and German as a control condition, but for these metaphors there are conventionally established translations which we have to use, but which may also introduce conceptual cross-linguistic differences into the material. Since a choice in translating the metaphor source and target concepts is ultimately a lexical choice, even when we pick the "best" translation, each lexeme may come with Fregean senses that could differ cross-linguistically. So while we aim to reduce conceptual cross-linguistic differences in the material to zero, it is unlikely we can fully achieve this given the practical constraints.

I take whatever unavoidable cross-linguistic sense differences that remain in the material to be crucial because they are of the sort that are classified as conceptual differences under CMT. If we simply set some arbitrary quality threshold, we would pretend that whatever conceptual cross-linguistic differences remain in those metaphors that we "pass" would be too small, too insignificant to affect reading/response times, plausibility judgements, or speakers' considerations of meaning salience that will be relevant in the context creation task. I firmly believe that we are not in a position, at least at the moment, to adequately judge where such a threshold should be set. So instead I propose we tally any and all deficits in the material in a way that it may inform our interpretation of the experimental results. I have already implicitly suggested what that form might be: probability. The size of the potential effect of non-conceptual aspects of idiomatic meanings that we will estimate in the experiments will be in the form of correlation coefficients. The linear mixed effects models we will use will tell us how much of the total variation in reading/response times and plausibility judgements is due to metaphors' language specificity. Hence, these effect size estimates are estimates of the prior probability that language-specific aspects of idiomatic meanings affect speakers' reading/response times and plausibility judgements. If we could make it so that there were no conceptual cross-linguistic

differences in the material, then the experimental effect size estimates of language specificity, if there is such an effect, would be entirely a function of non-conceptual cross-linguistic differences. If some deficits are unavoidable, then if we observed an experimental effect of language specificity, it would be a function of non-conceptual *and* a function of conceptual aspects of idiomatic meanings. If we can capture the deficits in the material as probabilities, we can use them to “correct” the experimental effect size estimates by deducting the proportion of conceptual deficits of the material. However, this requires that the substantive significance of both the experimental estimates and the deficits in the material are expressed in the same unit: probability.

Using the tools laid out in the previous chapter and the software G*Power for power analysis, we obtain the following tables. Tables (6.4) and (6.5) tally the deficits among the English-specific, German-specific, cross-linguistically shared, and novel metaphors. Average tendencies using sample size as weights are provided in the last row of each table. One of our quality standards is that if a certain amount of conceptual cross-linguistic differences in the form of Fregean senses is unavoidable in translating the material, these deficits should be not significantly bigger than the differences in the linguistic form of metaphors (e.g. differences in length which could affect reading/response times, or differences in corpus frequency which could be indicative of differences in familiarity). To this end, Table (6.6) lists the features of metaphors’ linguistic form and meaning ranked by the amount of cross-linguistic differences. Although Fregean sense differences rank high, aspects of metaphor length, which are much more likely to affect reading/response times, rank higher. Another central aspect of our quality standards is that whatever amount of cross-linguistic variation in the material is unavoidable, that variation should be similar for the four metaphor types so that experimental test and control conditions are equally affected. See Table (6.7) for a breakdown by linguistic form and meaning. Once we have gathered the various experimental effect size estimates, I will perform a meta-analysis on them to make the corrections necessitated by the deficits in the material. The metrics of statistical reliability provided in Table (6.4), (6.5), (6.6), and (6.7), will be essential in this meta-analysis.

6.4 Summary

In this chapter, we looked at the linguistic meaning of the metaphors that will be used in the experiments. I argued that they are cross-linguistically comparable between English and German and in that discussion paid special attention to the lexical meanings of metaphors’ source and target concepts, but crucially also to whether other idiomatic expressions and conceptual analogies are available which native speakers could use to comprehend language-specific metaphors from the language not familiar to them. The existence of these expressions and analogies, according to CMT, should suggest that speakers have conceptual metaphors that help them process complex linguistic metaphors and if speakers have the conceptual metaphors necessary to process particular L2 metaphors, these metaphors should be cross-linguistically intelligible to them.

I suggested that we compare the cross-linguistic similarity of metaphors’ source and target concepts by counting the number of senses for any Fregean sense that would have to be added, taken away, or exchanged for another sense in order to transform the English concept into its German counterpart. In order to quantify this I adapted the measure of Levenshtein

distance and I showed how this distance can be applied to multiple concepts since we have two per metaphor: a source concept and a target concept. Both of these distances, for lexical and compositional semantics, turned out to be quite comparable between English and German. And I showed that when senses were different between an English concept and its German counterpart, those senses were mostly irrelevant in the context of the metaphors that they appear in. The semantic Levenshtein distances and the contextual relevance of sense differences are expressed in such a way that we can interpret them as probabilities. Understood probabilistically, they represent the conceptual deficits in the experimental material which we will need in order to correct the effect size estimates of conceptual and non-conceptual aspects of metaphors' idiomatic meanings that we will measure experimentally.

Throughout this and the previous chapter I have made an effort to not just *argue* that the English and German translations of the metaphors used in the experiments are comparable between the two languages, but I have tried to *quantify* how comparable, how similar and dissimilar they are, which I have summarised in Table (6.3). The purpose of this table is that once we look at the experimental results, we can gauge if any of the aspects of linguistic form, familiarity, or meaning could be the source of rating differences in the plausibility judgement task. We can also see how the cross-linguistic dissimilarities of particular aspects of linguistic form, familiarity, and meaning might accumulate, if rating differences should be motivated by a set of these aspects rather than one of them in particular. Our main focus is on conceptual deficits in the material, but the table also allows us to ensure that non-conceptual deficits are also minimal and not more than the conceptual deficits. If non-conceptual deficits exist, we want them to be equally distributed across metaphor types, experimental conditions and manipulations so that we can interpret the results "all other things being equal."

What we will observe in the experiments is the difference in metaphor comprehension between L2 metaphors, on the one hand, and L1 and cross-linguistically shared metaphors, on the other, measured as differences in reading/response times and plausibility judgements. If CMT's assumption is correct, then monolingual native speakers should read L2 metaphors as fast as and judge them to be as intelligible as L1 and cross-linguistically shared metaphors. If, however, we find slower L2 reading times and lower L2 metaphor intelligibility, then this would contradict CMT's assumption, *but only if* it was not caused by conceptual deficits, deficits in selecting the metaphors in a way that maximises conceptual cross-linguistic similarity. So if there are conceptual deficits in the experimental material and we find slower reading times and plausibility judgements in the experiments, this could be caused by conceptual aspects of L2 metaphors' idiomatic meanings or by non-conceptual aspects or by a combination of both. So when we experimentally estimate the influence of conceptual and non-conceptual aspects of idiomatic meanings, we need to correct for deficits in the material by deducting the amount of conceptual deficits from the experimental estimate of the non-conceptual aspects and adding it to the estimate of the conceptual aspects. The conceptual deficits are the topic of this chapter.

Chapter 7

Experiments: cross-linguistic metaphor intelligibility out of context

7.1 Introduction

CMT proposes that complex concepts, such as the ones we find in linguistic metaphors, are composed from more basic concepts through a process that creates mappings (conceptual metaphors) between mental representations, topographic maps, schemas, and simpler concepts so as to yield ever more complex networks of mappings, called cascades, which correspond to complex inferences. If particular conceptual metaphors are used frequently, for instance, because of their commonness in language use, these conceptual mappings become more salient as inferential pathways. According to CMT, salient inferences such as the idiomatic meanings of metaphoric proverbs are thus evidence of salient ways of *conceptualising*, and, according to CMT, the systematicity of metaphors in linguistic expression gives insight into the kinds of conceptual metaphors. RT and GS, contrary to CMT, claim that while meaning salience may in part depend on the mental representation of concepts, the systematicity of metaphors in linguistic expression cannot be taken as *direct* evidence of conceptual metaphors and metaphor comprehension involves *non-conceptual* linguistic conventions which may take precedence over purely conceptual intuitions. In Section (4.2) in Chapter 4, I introduced the two central questions of this research, the first of which will be the subject of this chapter: Are metaphors from one language intelligible to speakers of another language if those two languages are in close linguistic contact and thus can be assumed to share similar or the same concepts?

We can delineate metaphors along two axes: meaning salience and the language specificity of their meanings. Metaphors with conventional idiomatic meanings have highly salient meanings. These conventional meanings, however, can be specific to a particular language, they can occur in several languages, or they can be universal. For the experimental material we therefore selected three types of metaphors with conventional idiomatic meanings: (1) metaphors that occur in both English and German, (2) metaphors that only occur in English, and (3) metaphors that only occur in German. Inferences drawn for novel metaphors are less salient than the conventional idiomatic meanings of metaphoric proverbs. Novel metaphors thus form a fourth

type of metaphor along the two axes: their meanings are nonsalient and cannot be language-specific because they have not been conventionalised. With respect to these metaphor types, we can define cross-linguistic metaphor communicability in two ways: L2 language-specific metaphors are *strongly communicative* if they are as communicative as L1-specific and cross-linguistically shared metaphors (shared between the L1 and L2) and L2-specific metaphors are *weakly communicative* if they are more communicative than novel metaphors. In the experiments of this chapter, cross-linguistic communicability of L2 language-specific metaphors is measured in three ways: (1) through reading/response times, (2) through plausibility judgements, and (3) through a context creation task. If CMT is correct, then L2 language-specific metaphors should be strongly communicative on all three measures.

If CMT is correct and metaphor comprehension is primarily a matter of conception, then when we pick a set of linguistic metaphors with salient conventional meanings such that the source and target concepts of these metaphors, their lexical meanings, their Fregean senses, expectations of contextual relevance, and related mental imagery is cross-linguistically approximately similar between two languages in close linguistic contact (i.e. maximised), then these language-specific metaphors should be equally communicative to speakers of either one of the languages. In Section (4.4.4) in Chapter 4, I suggested that languages in close contact exchange a lot of concepts, simple and complex ones, as well as the conceptual metaphors to construct complex mental imagery and I used book translations as evidence to argue that English and German have one of the closest language contacts in the world. In the previous two chapters, we saw that the linguistic form, familiarity of that form, and linguistic meaning of the metaphors in the experimental material is comparable, specifically, that most of the Fregean senses of the words in the metaphors are cross-linguistically approximately similar between the English and German version of the experiments. The senses that are different are almost always irrelevant in the context of the respective metaphors.

In the previous chapter, we saw that English and German native speakers can use similar conceptual analogies (*conceptual metaphors*) to come to the metaphoric meaning compositionally. Specifically, we saw how these conceptual analogies are cross-linguistically approximately similar for the cross-linguistically shared metaphors. For the language-specific metaphors, English and German speakers could use the conceptual metaphors they have in their native language in order to infer meanings for L2 language-specific metaphors that are similar to their idiomatic meanings. By selecting the material and the two languages in this way, it should be very likely that L2 language-specific metaphors will be cross-linguistically communicative and it is thus likely that CMT's predictions will be confirmed. If RT and GS are correct, however, then we should find evidence of *lowered* cross-linguistic communicability of L2 metaphors *despite* maximised cross-linguistic approximate conceptual similarity. What we find is that when monolingual native speakers of English and German see L2 metaphors out of context, they read, respond, and react more slowly and rate L2 metaphors to be less plausible than L1 or cross-linguistically shared metaphors. This is not the case for bilinguals.

7.2 Predictions

Predictions according to question 1 can be divided into two parts: (7.2) predictions concerning cross-linguistic intelligibility, and (7.2) predictions concerning potential differences between monolinguals and bilinguals.

Cross-linguistic intelligibility

CMT predicts: If intelligibility of metaphors requires conceptual metaphors and if the closer the linguistic contact between any two cultures is, the more they exchange concepts and align their inventories of conceptual metaphors (conceptual systems), then cross-linguistic intelligibility of metaphors is more likely between two languages in close linguistic contact.

The linguistic theories predict: The link between linguistic form and meaning is arbitrarily agreed upon by a speech community by means of conventions that are not necessarily conceptually more plausible than their conceptual alternatives, but are primarily motivated by a need to facilitate effective communication. The meaning of any linguistic symbol, such as a metaphor, does not have to be conceptually more sensible than its alternatives as long as it facilitates effective communication between speakers of a language. Metaphors are cross-linguistically intelligible between two languages if the metaphors happen to have the same arbitrary conventional meanings (sense relations) in those two languages.

For any two languages in close linguistic contact, let metaphor x be conceptually sensible for speakers of both languages but let it have different arbitrary conventional meanings (sense relations) in the two languages and let metaphor y have the same arbitrary conventional meaning in addition to being conceptually sensible in both languages. The linguistic theories would predict that metaphor y is more likely to be cross-linguistically intelligible between the two languages in close linguistic contact than metaphor x .

Monolinguals versus bilinguals

CMT predicts: Monolinguals will perceive metaphors' intelligibility as bilinguals do because for two languages in close linguistic contact both groups of speakers should have the same inventories of conceptual metaphors.

The linguistic theories predict: Monolinguals and bilinguals will perceive metaphors' intelligibility differently because even if for two languages in close linguistic contact the two groups of speakers agree on the metaphors conceptually, they will differ in their knowledge of metaphors' arbitrary linguistic meanings (sense relations).

7.3 Material and methods

7.3.1 Participants

In the experiments of this chapter participants see all metaphors out of context. Thirteen monolingual English native speakers (10 females; mean age = 19, SD = 1.2), 19 monolingual German native speakers (12 females; mean age = 41, SD = 16.2), and 60 English-German bilingual speakers (34 females; mean age = 39, SD = 15.4) took part in the out-of-context experiments. We want participants in the experiments to be familiar with the metaphors selected from their native language. I proposed that we rely on metaphors' frequency in corpora of English and German as an indicator of familiarity and selected metaphors accordingly. Linguistic corpora of written and spoken language represent a sample of the language that speakers have been and are exposed to. Hence, if we choose metaphors that are frequent in these corpora, then the likelihood that participants will be familiar with the metaphors is a function of the cumulative probability we calculated in Chapter 5. Through a diachronic corpus analysis, we

ensured that this probability has been stable throughout participants' lifetimes. The average participant is 41 years old or younger. Note that English monolinguals are generally younger and vary less in age than the German monolinguals and the bilinguals. This is because the English monolinguals are all university students whereas the German monolinguals and bilinguals include professionals as well as students.

Participants were asked to fill out a questionnaire where they provided the following information: age, gender, native language, additional languages, the native language of their parents. Participants were asked to answer the questionnaire after completing all experimental tasks so as not to bias them in their performance. German monolinguals and bilinguals had to answer additional questions: amount of second language education, amount of time spent abroad in an English-speaking country (monolinguals, less than 6 months spent in an English-speaking country (most stated typical two-week vacations which accumulated to the average) and less than a year of English as a second language; more than half of the German monolinguals had not had English as a second language at all because they live in East Germany where they learned Russian as a second language at school; bilinguals, few natural bilinguals, most advanced second language learners with an average of 6 years of English as a second language in secondary school and 5 years at university and an extended stay in an English-speaking country of more than 2 years, on average, and at least one year of that spent continuously in an English-speaking country).

For the purpose of the experiments, *monolinguality* is defined as follows: Speakers are monolingual if they are L1 native speakers and they do not speak the L2 or well below a level of fluency necessary for the metaphors in the experimental material. *Bilinguality* is defined as follows: In the experiments, almost all bilinguals are advanced L2 learners (near natives); a few are natural bilinguals where one parent or both parents are native speakers of the L2. In order to categorise participants in the experiments accordingly, language proficiency was recorded as the number of years of institutional instruction in the non-native language, the duration of time spent abroad, language proficiency of parents and frequency of L2 use. Bilinguals include advanced second language learners with an extended stay or permanent residence in the L2 country.

Bilinguals serve as a control group because they are familiar with both the language-specific and cross-linguistically shared metaphors' idiomatic meanings, whereas monolinguals will only be familiar with the idiomatic meanings of L1 and cross-linguistically shared metaphors. Bilinguals should therefore find the language-specific metaphors to be conceptually as plausible as the shared metaphors. If monolinguals should find L2 metaphors to be conceptually less plausible than shared metaphors, then we can compare the magnitude of this diminished plausibility to the bilinguals, making them a control group. We therefore have three control groups: monolinguals' ratings for the shared metaphors in the judgement task (strong intelligibility), their ratings for the novel metaphors (weak intelligibility), and bilinguals' ratings in the judgement task for language-specific and shared metaphors.

7.3.2 Material

There are 12 German-specific metaphors, but only 5 English-specific metaphors. The reason for this is that 6 of the 12 German-specific metaphors were originally used in a pilot study during a training phase preceding the actual judgement task where participants familiarised

themselves with the rating scale. Later in the actual experiments, the rating scale was displayed at the bottom of each metaphor slide. The training phase was thus no longer necessary. The English metaphors had already been verified through the extensive corpus analysis and so the 6 German metaphors but no additional English metaphors were included in the test phase. There are 13 cross-linguistically shared metaphors and 11 novel metaphors. Their numbers are thus comparable to the 12 German metaphors. This brings the total number of metaphors to 41. See Table (A.1) in the Appendix.

7.3.3 Procedure

Experiments were conducted both in the lab and via the internet. I will describe their experimental setup in turn and then show their results are qualitatively and quantitatively comparable.

Experiments conducted in the lab were set up in the software E-Prime. I will describe all of the instructions and the way that the metaphors were presented in turn. Experiments were set up at a screen resolution of 1280×768 , 16 bit colour depth, and 60 Hz refresh rate, with white text on black background. First, participants saw a 'welcome screen', followed by these instructions (English monolinguals saw them in English, German monolinguals and English-German bilinguals saw them in German):

In this experiment, you will see sentences such as "Time is money."	In diesem Experiment sollen Sie Sätze vom Format "Zeit ist Geld" lesen und bewerten.
Your task is to rate how much sense a sentence such as "Time is money" makes.	Ihre Aufgabe ist es, zu bewerten wieviel Sinn jeder Satz für Sie macht.
To rate a sentence press numbers 1 through 5. Press 1 if you think the sentence "makes no sense" and 5 if you think it "makes sense."	Nutzen Sie bitte die Tasten 1 bis 5 um die Sätze zu bewerten. Dabei steht 1 für Sätze, die "keinen Sinn machen" und 5 für Sätze, die "Sinn machen."
But first, let's practice.	Aber zunächst einige Übungssätze.
When you are ready, press the spacebar to begin the training.	Wenn Sie bereit sind, drücken Sie die Leertaste.

The phrasing of the task is deliberately left open-ended. We do not want to confuse participants too much with lengthy instructions about what they should and should not do because (1) when participants are told not to make judgements in a certain way and that conflicts with their natural intuition, they will either consciously ignore the instructions or unconsciously revert back to their natural intuition anyway, and (2) we control how participants can approach the task by controlling the information available to them to draw inferences, i.e. by controlling the independent variables we are manipulating and by having ensured the familiarity of the metaphors, and by having an understanding of what inferences may arise by knowing about the metaphors' lexical and compositional semantics as well as their relevance expectations. The phrasing of the task "how much sense does a sentence [i.e. metaphor] make" is identical between English and German, in terms of its linguistic form as well as all of its semantics including connotations and sociolinguistic register of the idiomatic expression. This broad, catch-all

phrasing specifically elicits a layman's notion of *sense* (not to be confused with the linguistic notion of sense relations) that is colloquially synonymous with *speaker meaning* which is (a) the notion of meaning we intend to test and (b) using the layman's term *sense* avoids linguistic terminology such as *semantics*, *pragmatics*, *speaker meaning*, *not the utterance meaning*, *not the literal meaning*, and even the notion of a *conventional meaning* because all of these have various interpretations for those participants who have been exposed to linguistics and those who are non-linguists. I therefore thought it best to avoid any and all of these linguistic terms.

Next, participants practised the use of the ratings scale on six metaphors. The presentation of the practice metaphors was the same one used in the main trial: metaphors were presented one at a time. Each metaphor was centred on the screen, beneath it the rating scale (1 = *makes no sense* to 5 = *makes sense*). In between metaphor slides, +-----+ was displayed as fixation with a white font on black background, centred on the screen. The order of metaphor slides was random within the list of practice metaphors and within the list of trial metaphors. After participants had rated all practice metaphors they saw the following instruction that reminded them of their task again. English monolinguals saw the instructions in English, German monolinguals and bilinguals in German:

You have completed the training.	Sie haben erfolgreich alle Übungssätze bewertet.
You are now ready to proceed to the actual experiment.	Sie sind nun bereit mit dem eigentlichen Experiment zu beginnen.
Remember, rate how much sense a given sentence makes to you. Press 1 (makes no sense) through 5 (makes sense).	Denken Sie daran, zu bewerten, wieviel Sinn die Sätze für Sie machen und nutzen Sie dazu die Tasten 1 (macht keinen Sinn) bis 5 (macht Sinn).
When you are ready, press the spacebar to begin the experiment.	Wenn Sie bereit sind, drücken sie die Leertaste um das Experiment zu beginnen.

At the end they saw a 'goodbye screen' that thanked them for their participation. After that, participants went on to part 2 of the experiment, the metaphor context creation task. The instructions were as follows (English monolinguals saw them in English, German monolinguals and bilinguals in German):

Please think up continuations that would make the given sentences more sensible. All your continuations should start with a preposition (e.g. on, to, of, for, etc.), but you can add whatever you want. But, please, keep your continuations as short as possible.	Denk dir bitte Fortführungen aus, die die folgenden Sätze sinniger machen. Alle Fortführungen sollten mit einer Preposition beginnen (z.B. auf, zu, von, mit), aber danach kann folgen was will. Aber bitte versuche dich kurzzufassen.
---	---

Consider the following examples:	Zum Beispiel:
<i>Beauty is a ticket</i> might not make much sense, but adding <i>to fortune</i> could make it more sensible.	<i>Schönheit ist eine Eintrittskarte</i> allein macht vielleicht weniger Sinn als wenn man <i>zum Ruhm</i> hinzufügt.

Similarly *dreams are froth on an ocean* might be more sensible than *dreams are froth*.

In ähnlicher Weise ist *Träume sind Schäume auf dem Ozean der Nacht* vielleicht sinniger als nur *Träume sind Schäume*.

All speakers (English monolinguals, German monolinguals, and bilinguals) were given the same 8 metaphors translated into English and German: *ignorance is bliss* (cross-linguistically shared metaphor), *silence is golden* (cross-linguistically shared), *anger is a volcano* (cross-linguistically shared), *love is a journey* (English-specific), *sports is murder* (German-specific), *the devil is a squirrel* (German-specific), *a home is money* (novel), *a friend is a magnet* (novel). English monolinguals created continuations in English, German monolinguals and bilinguals in German.

For experiments conducted via the internet, part 1 had the same division into practice phase and trial phase as the experiments conducted in the lab, with the same instructions. Metaphors were rated in sets of ten and on the same scale as in the lab, a 5-point Likert scale. Part 2 had the same instructions used for experiments in the lab; participants' responses in part 2 were collected directly after part 1, i.e. they had to complete both parts in order to complete the online experiment. Only after they had completed part 2 were they asked to fill out a form with personal information with the same questions as those tested in the lab. At the end, they saw a 'thank you for participation' and 'goodbye screen.'

7.3.4 Design

Outcome variables are reading/response times measured in milliseconds and plausibility judgements rated on a 5-point Likert scale. Predictor variables are (1) metaphor type (English-specific metaphors, German-specific metaphors, cross-linguistically shared metaphors that occur in both English and German, and novel metaphors that occur neither in English nor German), (2) speaker group (English native speakers who do not speak German = English "monolinguals," German native speakers who do not speak English = German "monolinguals," and English-German bilingual speakers who are born native speakers of German, are fluent in English, and reside in an English-speaking country), (3) age, (4) gender, and (5) metaphors' corpus frequency.

We would expect that some participants respond more quickly than others, in general, and we would expect that some participants might tend to avoid giving the highest or lowest plausibility ratings. At the same time we would expect that metaphors of the same type will not all be read and responded to equally quickly and will not uniformly receive similar ratings. Since participants' average reading/response speed and the variation in reading/response times of metaphors of the same type are not independent of each other and since participants' use of the rating scale and the variation in plausibility ratings given to metaphors of the same type are not independent of each other, linear mixed effects models are used in all analyses.

7.4 Results

Cross-linguistic metaphor communicability and intelligibility is measured in three ways in the experiments: (1) through reading/response times, (2) through plausibility judgements, and (3) through a context creation task. Reading/response times and plausibility judgements are quantitative measures, while the context creation task gives qualitative results. Therefore, the

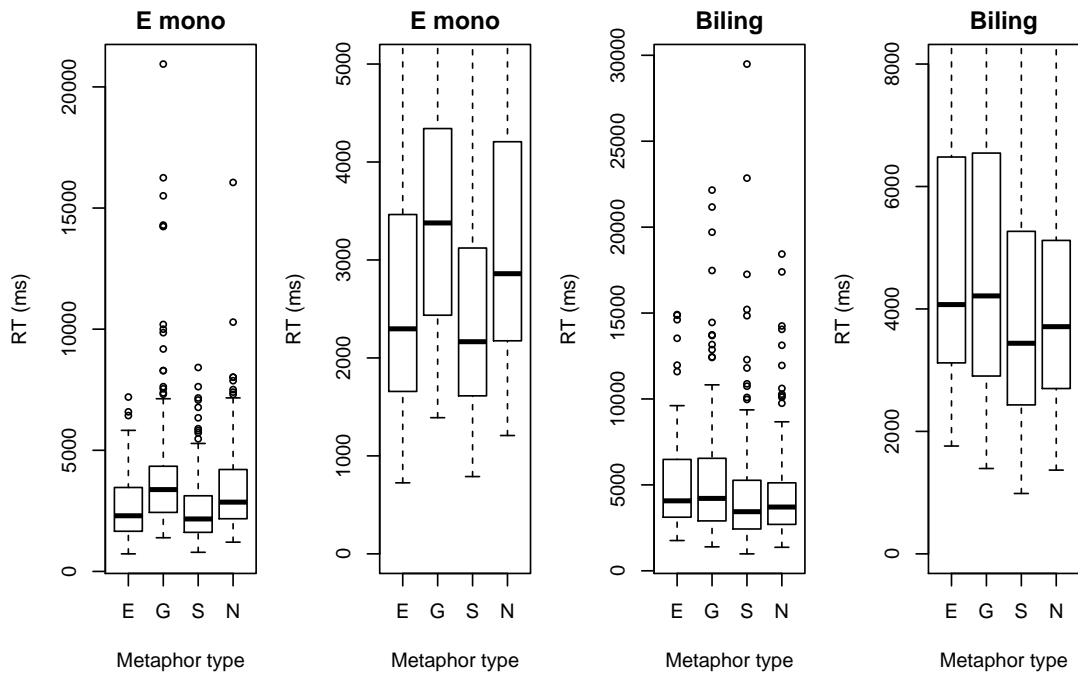


Figure 7.1: Reading/response times (RT) in milliseconds (ms) for English monolinguals (E mono) and English-German bilinguals (Biling) and for each metaphor type: English-specific (E), German-specific (G), cross-linguistically shared (S), and novel (N). The second and fourth graph is the first and third zoomed-in on faster RTs.

quantitative results will be discussed in this section, Section (7.4), and the qualitative results of the context creation task will be discussed in Section (7.4.4).

7.4.1 Reading/response times

Initially I intended to use reading/response times only to check that speakers would respond similarly quickly. My main focus was on plausibility judgements. This was an initial concern because participants were not put under pressure to respond quickly. Some participants took more than 8 seconds for particular metaphors. In the data analysis, these extreme reading/response times are of course excluded. My concern was focussed on the individual speaker and it was not until I redid the data analysis with linear mixed effects models that it occurred to me to try checking reading/response times on the level of speaker groups because the ANOVA used in the initial analysis revealed *no* effect of metaphor type on reading/response times per speaker group. By the time of the re-analysis using mixed models, plausibility judgements had already been collected for out-of-context and in-context experiments. Collecting reading/response times in the in-context experiments would have meant recruiting a completely new batch of participants, at least another 60 to 70 (judging from power analysis on the results so far). For German monolinguals, reading/response times were not collected because experiments were conducted via the internet where measurement accuracy could not be ensured. It would thus definitely be valuable to redo the experiments with reading/response times. Since this only became apparent after the main analysis of the data, this thesis has to limit itself to reporting the results thus far.

All experiments, for which reading/response times are reported in this thesis, were conducted in the lab and with the software E-Prime which ensures adequate measurement accuracy.

Recall that the idiomatic meaning of metaphoric proverbs can be understood as an inferential shortcut: instead of having to consider all conceptually possible alternative interpretations and assessing their plausibility in the context of the analogy implied by the metaphor, speakers who are familiar with the idiomatic meaning are provided with a default interpretation, which eliminates the need to consider all alternatives. CMT claims that idiomatic meanings of language-specific metaphoric proverbs are language-specific conceptual metaphors that together derive the complex concept which corresponds to the idiomatic meaning from basic primary embodied concepts. RT and GS claim that language-specific idiomatic meanings are not necessarily only language-specific ways of *conceptualising*, but can also include *non-conceptual* knowledge provided through linguistic convention. If CMT's claim regarding the nature of metaphoric idiomatic meanings is correct, then for two languages such as English and German and metaphors such as the language-specific metaphoric proverbs, where the metaphors' source and target concepts, expectations of contextual relevance, and associations with other mental imagery are cross-linguistically approximately similar, native speakers of English, for instance, should be able to infer linguistic meanings for language-specific German proverbs that are similar to the idiomatic meanings that German native speakers know. Hence, if CMT's claim is correct, we should not expect English speakers to show significant delays in processing for language-specific German metaphors. If RT and GS are correct, then we should expect to see such delays and the magnitude of the delays should correlate with the cognitive effort afforded to consider the plausibility of conceptual alternatives: English monolinguals' reading/response times for German-specific metaphors should be significantly slower than their reading/response times for English-specific and cross-linguistically shared metaphors. Bilinguals serve as a control group because, contrary to monolingual native speakers of either English or German, they should be familiar with the idiomatic meanings of both English and German language-specific proverbs as well as all cross-linguistically shared proverbs, and they should therefore be able to use inferential shortcuts for all language-specific metaphors just as for cross-linguistically shared ones.

It is striking that German metaphors attracted such slow responses among the bilinguals, given that the bilinguals are German native speakers and that the language of presentation was German. Notice that bilinguals' responses to English metaphors are similarly slow and slower than for cross-linguistically shared metaphors. One potential explanation could therefore be that bilinguals show slower responses for language-specific metaphors than non-language-specific cross-linguistically shared metaphors because they are sensitive to metaphors' language specificity. However, more experiments with bilinguals are needed to shed more light on this issue.

Visual inspection of Figure (7.1) reveals (1) a tendency for English monolinguals (E mono, in Figure 7.1) to read and respond slower for unfamiliar metaphors than for familiar ones and a tendency for English-German bilinguals (Biling) to read and respond slower for language-specific metaphors than for non-specific ones, (2) lots of outliers, i.e. participants whose reading/response times are much slower than the average participant's (marked with circles in Figure 7.1), and (3) median reading/response times, which in light of the large number of outliers are a better measure of average reading/response times than the mean, are generally low (~ 4 seconds or less, see also Table 7.1), which suggest that participants' responses are

Table 7.1: Median reading/response times (MED, and standard error, SE) for English monolinguals and English-German bilinguals.

RTs	E mono		Biling	
	MED	SE	MED	SE
E	2297 ± 175 ms		4072 ± 433 ms	
G	3378 ± 225 ms		4213 ± 320 ms	
S	2166 ± 102 ms		3440 ± 297 ms	
N	2860 ± 160 ms		3711 ± 261 ms	

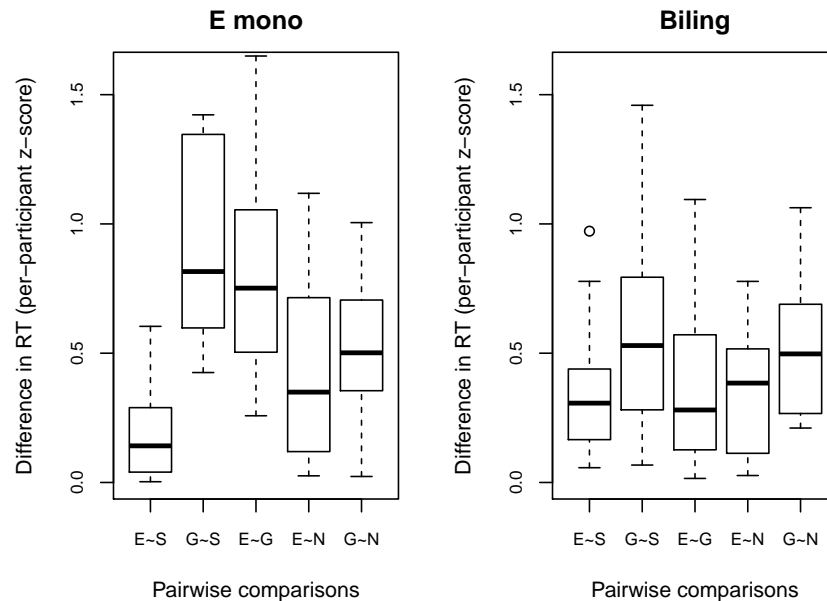


Figure 7.2: Difference in reading/response times (RTs) where RTs were converted to z-scores for each experimental participant individually. Boxes and whiskers indicate between-participant variation in z-scores.

quite intuitive. Recall that plausibility judgements and reading/response times were recorded simultaneously. It is therefore reasonable to assume that plausibility judgements (see Section 7.4.2) are also quite intuitive. The median reading/response times given per metaphor type in Figure (7.1) are averaged across participants. The large number of outliers (about half of participants) suggests that averaging across participants is obscuring potential per-participant reading/response time tendencies.

We would expect that some participants read and respond more slowly than others, in general. We can normalise reading/response times by converting them to z-scores

$$z = \frac{x - \mu}{\sigma}, \quad (7.1)$$

where μ is an experiment's sample mean, σ is an experiment's sample standard deviation in responses, and x is any measured reading/response time that we wish to convert to a z-score. In our case, z-scores are comparable across experiments because the out-of-context

reading/response times are comparable in terms of the estimated population mean and standard error. Average reading/response times are 3288 milliseconds for English monolinguals and 3859 milliseconds for bilinguals. Standard error is 2252 milliseconds for English native speakers and 2312 milliseconds for bilinguals. Normalisation using z-scores should thus be adequate. Alternatively, t-values are reported for the slopes and fixed effects in Tables (7.3), (7.5), and (8.2).

Shown in Figure (7.2) are reading/response time differences where reading/response times were converted to z-scores per participants, i.e. that participant's average reading/response time and deviation. $E \sim S$ is the absolute difference in reading/response time between the reading/response times for English metaphors (E) and reading/response times for cross-linguistically shared metaphors (S). Similarly, $G \sim S$ is the difference between reading/response times for German metaphors (G) and reading/response times for shared metaphors (S), and $E \sim G$ is the difference between reading/response times for English metaphors (E) and reading/response times for German metaphors (G). We see that English monolinguals read and respond similarly quickly for English and cross-linguistically shared metaphors (0.89 and 0.81 standard deviations) and 0.68 standard deviations slower for German metaphors than for English and shared metaphors $((0.89 + 0.81)/2 - 0.17)$. Bilinguals read and respond similarly quickly for English and German metaphors (0.36 and 0.38 standard deviations), but not as quickly as for cross-linguistically shared metaphors. The picture we see therefore resembles the predictions made by RT and GS, but not those made by CMT. Monolingual native speakers read and respond slower to German metaphors than to English or cross-linguistically shared metaphors, but bilinguals did not.

For monolinguals, the reading/response time differences calculated as z-scores for each participant individually show a similar picture to the median reading/response time per metaphor type (but averaged across participants), meaning that although monolinguals vary in their mean reading/response time, the tendency to read and respond to German metaphors more slowly than for English and shared metaphors is present for each of them. For bilinguals, reading/response time differences calculated on a per-participant basis show quite a different picture from the median reading/response time that averages across participants, meaning that individual bilinguals have different tendencies whether they read and respond to English or German metaphors more slowly than shared ones such that when we average across participants and look at all bilinguals collectively, these tendencies annihilate each other. In this case, the reading/response time differences calculated as z-scores on a per-participant basis are more insightful and reliable than the average.

The reading/response times of individual bilinguals differ 0.22 standard deviations more between German and shared metaphors than between English and shared or between German and English metaphors $(0.59 - (0.38 + 0.36)/2)$, whereas for the average bilingual, in Figure (7.1), it seemed that the difference between reading/response times for English and German metaphors should be smallest and the reading/response time difference between English and German metaphors should both be similarly different from the reading/response times of shared metaphors. Hence, on the level of the individual bilingual, German metaphors tend to have the slowest reading/response times, shared metaphors have the fastest reading/response times, and the reading/response times of English metaphors tend to be slower than for shared ones but not as slow as for German metaphors. More importantly, however, notice that this tendency is much less pronounced for bilinguals (0.22 standard deviations) than English monolinguals'

tendency to have slower reading/response times for German metaphors (0.68 standard deviations); monolinguals' tendencies are 3 times more pronounced than bilinguals' tendencies. Also notice that while differences in reading/response times between English and shared metaphors are quite small across monolingual speakers (0.17 standard deviations), reading/response time differences are generally higher across bilinguals (0.36 and 0.38 standard deviations). Hence, inter-personal variation in reading/response time differences is greater for bilinguals than for monolinguals. Note, however, that these differences are not statistically tested (nor is it easy to do so).

What we want to know is whether reading/response times for L2 language-specific German metaphors are significantly slower if we consider both by-item and by-subject variation *simultaneously*. A linear mixed effects model allows us to do just that. The models fitted to the reading/response times of English monolinguals and English-German bilinguals use the following formula:

$$y = X\beta + Z_1b_1 + Z_2b_2 + \varepsilon \quad (7.2)$$

where y is a vector of observed reading/response times, with mean $\mu(y) = X\beta$. A linear mixed effects model, or multiple regression, differs from a linear regression in that a linear regression only has fixed effects and a random error term ε . A mixed effects model has both fixed and random effects. The random effects of a mixed effects model thus explicate and give more insight into the random error term of a fixed-effects-only model. In the linear mixed effects model, β is the fixed effect, metaphor type, and b_1 and b_2 are random effects; b_1 is by-subject variation in reading/response times y and b_2 is by-item variation in reading/response times y . X , Z_1 , and Z_2 are model matrices of regressors relating the observed reading/response times y to the fixed effect β and the random effects b_1 and b_2 . Matrix X contains random intercepts; Z_1 and Z_2 contain random intercepts *and* random slopes. As in a fixed-effects-only model, in a mixed effects model ε is a vector of random error terms. It is the proportion of variance in reading/response times y that is not accounted for by the fixed and random effects. Since only the fixed effect, metaphor type, and the random effects, by-subject and by-item variation, are explicitly modelled, ε is implicitly given by the residual of the fitted linear mixed effects model, i.e. the variance in reading/response times y that is left “unexplained” by the fitted model. Thus the proportion of the variance in reading/response times y explained by the fixed and random effects and the “unexplained” variance ε tells us how well a linear mixed effects model fits observation y . Much of what I say about the reasons for using linear mixed effects models for the reading/response times directly applies to the mixed effects models of plausibility judgements because the models for plausibility judgements use the same model formula. The only difference is that instead of regressing to reading/response times, the models regress to plausibility ratings.

Assumptions of linear mixed effects models

See Locker, Hoffman & Bovaird (2007) and Bolker, Brooks, Clark, Geange, Poulsen, Stevens & White (2009) for an authoritative source on the assumptions of linear mixed effects models. Every participant rates multiple metaphors of each type instead of just one English-specific, one German-specific, one cross-linguistically shared, and one novel metaphor. Some participants might, for instance, regardless of metaphor type, be generally faster at reading and rating metaphors than other participants. Multiple reading/response times per participant are thus

not independent of each other. If metaphor type has a systematic effect on reading/response times, then when there are multiple metaphors per metaphor type, we would expect their reading/response times to be similar. At the same time, we would expect that even if metaphor type has an effect, that effect might affect some participants more than others. Nevertheless, regardless of the per-subject magnitude of the effect of metaphor type, variation in reading/response times should be systematic for all participants. Therefore, the reading/response times of multiple metaphors per metaphor type are not independent of each other. So there are two sources of non-independence of data points: by-subject and by-item variation in reading/response times, which could overshadow the effect of metaphor type on reading/response times that we wish to study. The linear mixed effects models used in the analysis resolve these two non-independences of data points by including by-subject and by-item variation relative to metaphor type as random effects: Z_1b_1 and Z_2b_2 , respectively. Z_1b_1 calculates different (random) intercepts and (random) slopes per reading/response time for every subject and every metaphor type. In other words, the model expects different baseline reading/response times (the intercepts) and effects of metaphor type on these ratings (the slopes) per subject. Z_2b_2 means that the model expects different baseline reading/response times and effects of metaphor type on these reading/response times per metaphor item within each metaphor type. The linear mixed effects models therefore resolve the two non-independences that would be problematic in linear models that only have fixed effects but no random effects.

A multiple regression fits linear regressions to the data. This presupposes that the data points follow a linear function, of course. The assumption is met if whatever the effect of metaphor type on reading/response times and plausibility judgements is, that effect changes reading/response times and plausibility ratings in a linear fashion. This assumption is satisfied, as we shall see in the analysis. One way to test this assumption is to look at the residual plots, i.e. the variance of data points about the fitted regression line. Figure (A.4) in the Appendix shows four residual plots, two for the linear mixed effects models fitted to the raw reading/response times of monolinguals and bilinguals, the other two where only reading/response times < 8000 ms were included in the models. There is no obvious pattern in the variance of data points. A linear regression is therefore appropriate and the variance is homoscedastic (randomly distributed about the regression line). The residual plots in Figure (A.4) also show that for both English monolinguals and bilinguals, the variance in reading/response time is not homoscedastic when outliers with reading/response times over 8000 ms are included in the models: higher fitted values have larger residuals than smaller values; the models are thus less “good” predictors for larger mean reading/response times, which, however, is what we would expect as larger reading/response times correspond to outliers. For the range we are most interested in and to which the mixed models were fitted, reading/response times faster than 8000 ms, the variance *is* homoscedastic. Excluding outliers thus not only gives a better fit of the linear mixed effects models overall, but also meets the homoscedasticity requirement.

Another assumption is that the variance of data points about the regression line should not only be homoscedastic but the density of their variance should be normally distributed, i.e. data points should be denser the closer they are to the regression line. The histograms in Figure (A.5) give the probability densities for English monolinguals and English-German bilinguals. The histograms in the left column are the raw residuals; in the right column, reading/response times over 8000 ms have been excluded. The actual kernel density of reading/response times is plotted as a continuous line. The corresponding normal distribution is plotted as a dashed line.

The actual residuals (as given in Figure A.4) are plotted as a rug along the x-axis. By comparing the kernel densities to the normal densities we can see that in the raw data bilinguals' reading/response times are not normally distributed, while monolinguals' reading/response times are. When reading/response times over 8000 ms are excluded, however, the kernel densities of both monolinguals and bilinguals are normally distributed along the regression of the corresponding linear mixed effects models.

Another assumption of linear models is that cases of collinearity should be avoided. Collinearity occurs when the linear behaviour of one fixed effect depends on the linear behaviour of another fixed effect. The model cannot fit two regressions such that one depends on the other but will try to find an average linear regression. The assumption of non-collinearity can only be violated when there are multiple fixed effects in the model. There is only one fixed effect in all linear mixed effects models of reading/response times and plausibility judgements, in this chapter and the next, which is metaphor type. Therefore, collinearity cannot be an issue and the assumption of the model is satisfied.

The basic assumptions of linear mixed effects models are thus met and it is therefore adequate to use them in an analysis of the reading/response times. The three main questions in an analysis using linear mixed effects models are: (1) How well do the models fit the observed variation in reading/response times? (2) How much do reading/response times vary because of the fixed effect, metaphor type, and does metaphor type have a greater effect than by-subject and by-item variation? (3) How exactly does metaphor type affect reading/response times and is that effect significant?

Model fit and effect sizes

The linear mixed effects models were fitted using the lme4 package (Bates, Maechler, Bolker & Walker, 2014) in R (R Core Team, 2013) and REML (Restricted Maximum Likelihood) was used as a fitting method, which gives better fit than the alternative ML method. The variation not accounted for by the fixed and random effects in the model is the random error term ε . The more ε converges toward zero, the more comprehensive the model is, i.e. the better its fit and the greater its explanatory power. If the amount of "unexplained" variation is smaller than the variation of the random effects, then it is unlikely that we are missing a factor. If we were missing a factor, the amount of "unexplained" variation should be at least as big as that of the random effects and probably be somewhere between that of the random effects and the fixed effect. The ratio between the variation explained by the fixed effect, metaphor type, and the random effects, by-subject and by-item variation, should give us an idea of the size of the effect of metaphor type on reading/response times and plausibility judgements.

R^2 is a popular estimator of variation explained by linear fixed-effects-only models. Ronghui Xu (2003) recommends Ω^2 as an estimator of variation explained by linear mixed effects models and shows that it gives estimates comparable to R^2 in linear regression. Ω^2 , also written as R_X^2 (X for Ronghui Xu) by others in the literature, takes the residual sums of squares as a measure of variance. As of writing, there is no general agreement which method gives the best R^2 estimates for mixed effects models. In this and the next chapter, I use R_X^2 as an estimate of R^2 because it is currently among the better methods and the one that should be most intuitive for readers familiar with R^2 in linear regression.

The linear mixed effects model fitted to the reading/response times of English monolinguals

Table 7.2: Variance explained by each fixed and random effect in the linear mixed effects models: reading/response time models. Each value is a ratio of the variance accounted for by that effect and the total model-wide variance. These ratios are effectively R^2 estimates ranging from 0 to 1 (for ease of comparison, given here as percentages). The higher the value, the more of the overall variance is explained by the respective effect.

Effect	E mono	Biling
Metaphor type	39.4%	44.5%
By-subject	23.5%	33.6%
By-item	7.7%	7.0%
Error term	29.4%	14.9%

fits 70.6% (see Table 7.2, 39.4% + 23.5% + 7.7%) of the total RT variance observed in the data. The model fitted to the reading/response times of English-German bilinguals accounts for 85.1% (44.5% + 33.6% + 7.0%) of the total variance observed in reading/response times. The effect we are interested in, the fixed effect *metaphor type*, accounts for more of mono-linguals' total variance in reading/response times than either by-subject or by-item variation (39.4% > 23.5% and 39.4% > 7.7%) and more than by-subject and by-item variation taken together (39.4% > 23.5% + 7.7%). The same is true for the linear mixed effects model fitted to bilinguals' reading/response times: metaphor type accounts for more of the total variation in reading/response times than either by-subject or by-item variation (44.5% > 33.6% and 44.5% > 7.0%) and more than both random effects taken together (44.5% > 33.6% + 7.0%). Hence, most of the variation in reading/response times that the models fit so well is due to the fixed effect *metaphor type* and not due to the random effects.

Statistical significance of the effect of metaphor type on reading response times

The full model, repeated in Formula (7.3), estimates the effect of metaphor type on reading/response times, with per-subject and per-item variation in reading/response times as random effects, by estimating how well metaphor type predicts variation in reading/response times.

$$y = X\beta + Z_1b_1 + Z_2b_2 + \varepsilon \quad (7.3)$$

$$y_0 = Z_1b_1 + Z_2b_2 + \varepsilon \quad (7.4)$$

The null model, shown in Formula (7.4) and denoted by the index zero for observed reading/response times y , is the full model *without* the variable whose effect on reading/response times we are interested in, *metaphor type*. It otherwise contains the same random effects, and random intercepts and slopes for per-subject and per-item variation in reading/response times. I am following the recommendation by Barr, Levy, Scheepers & Tilly (2013) and Schielzeth & Forstmeier (2009) here. They argue that models such as these that contain random slopes and random intercepts for random effects in both the full and the null model are less prone to Type I errors than intercept-only models. Constructed in this way, the null model is thus “unaware” what type a particular metaphor is.

The data we gather for random effects should exhaust the population of interest and the

levels of the effects. We are investigating metaphor comprehension in English and German, so in order to exhaustively sample by-subject variation we should include both speakers of English and German and both monolingual native speakers and bilingual speakers. This is the case. We have therefore exhausted the levels of the random effect by-subject variation. In Chapter 3, I suggested that the best way to think about the four metaphor types that form the material of the experiments is as them being delineated along a scale in terms of their *meaning salience* (Giora, 1997, 1998, 1999, 2003; Peleg et al., 2008; Peleg & Giora, 2011): L1 language-specific and cross-linguistically shared metaphors have conventional idiomatic meanings that should be highly familiar and thus salient to monolingual native speakers. The novel metaphors have no conventionally established idiomatic meanings and so their potential meanings are less salient than the idiomatic meanings of conventional metaphors. The idiomatic meanings of L2 language-specific metaphors should be unfamiliar to monolingual speakers and so the meanings that these speakers do infer should be less salient than the idiomatic meanings of metaphors they know (L1-specific and cross-linguistically shared ones). Bilinguals, however, should be familiar with the idiomatic meanings of L2 metaphors and their meanings should thus be more salient to these speakers than to monolinguals. The four metaphor types thus allow us to look at the conventionality and language specificity of idiomatic meanings at the same time.

In Chapter 3, I also said that the meanings of “dead” metaphors, although highly conventional and salient, cannot be considered as “idiomatic” because in order for a metaphor’s meaning to be idiomatic, non-idiomatic (“literal”) interpretations need to be inferential alternative outputs. This is not the case for “dead” metaphors; their conventional meanings are non-idiomatic. Sampling items from the four metaphor types thus exhausts the levels and population of by-item variation. By-subject and by-item variation in reading/response times, as modelled in the linear mixed effects models, are thus adequately represented as random effects in the models. While the experimental data gathered for random effects should *exhaust* the population of interest and the levels of the effect, fixed effects usually only *sample* from the population of interest. The levels of the fixed effect, the types of metaphors, as we saw, exhaust the degrees of meaning salience, however, because there are many more metaphors of each type whose reading/response times we could have sampled, we have only sampled the population of interest for metaphor type and not exhausted it. It is thus adequate to model metaphor type as a fixed effect in the linear mixed effects models.

Because the random effects exhaust the population of interest and the levels of the effects, it is therefore adequate to treat the range of reading/response times and plausibility judgements as representative of population-wide variation and a model constructed from them that is “unaware” of a metaphor’s type is an adequate null model. Comparing a model that is “aware” of metaphors’ type against such a null model is then equivalent to gauging whether differences in reading/response times or plausibility judgements between metaphor types in the full model could be coincidental or not. In other words, if the null model is a representation of the variation in reading/response times or plausibility judgements that we would normally expect when metaphor type plays no role, then comparing the full model, which distinguishes between metaphor types, against the null model is equivalent to comparing variation between metaphor types against their normally expected distribution. Hence, in our particular case the p-values given by comparing the full linear mixed effects models to the null models through likelihood ratios are directly comparable to p-values given by ANOVAs—only the mixed effects models consider both by-subject and by-item variation simultaneously while ANOVAs do not.

Table 7.3: Out-of-context reading/response times in milliseconds. Best-fitting regression using mixed-effect modelling. Models fitted to responses of English monolingual native speakers (E mono, $R^2 = .706$) and English-German bilinguals (Biling, $R^2 = .851$). Intercept for metaphor type is cross-linguistically shared metaphors. P-values are likelihood ratios (χ^2 tests) between the linear mixed effects model that includes the variable of interest, *metaphor type*, and the model that does not (the null model). Significance codes: *** for $p \leq .001$, ** for $p \leq .01$, * for $p \leq .05$, † for $p \leq .1$.

Speaker	Predictor	B	St err	<i>t</i>	<i>p</i>
E mono	S (intercept)	2531.2	244.0	10.374	< .001 ***
	E (L1)	178.8	203.6	.878	.3596
	G (L2)	948.1	172.7	5.490	.0000082 ***
	N	775.0	216.9	3.573	.0011 **
Biling	S (intercept)	3682.2	325.4	11.316	.2768
	E (L2)	571.6	340.4	1.679	.0915 †
	G (L1)	338.3	313.4	1.080	.2849
	N	160.1	267.1	.599	.5506

Wilk’s theorem (1938) states that, given a sufficiently large number of samples, the likelihood ratio between two nested models is asymptotically χ^2 -distributed. The χ^2 value that the likelihood ratio approximates can then be used to find the corresponding statistical significance. The degrees of freedom of this approximate statistical test is the difference of the number of degrees of freedom of the full model minus the number of degrees of freedom of the null model. In our case, the full model is “aware” of the 4 metaphor types (English-specific, German-specific, cross-linguistically shared, and novel metaphors). The null model is “unaware” of the 4 metaphor types and thus treats all metaphors as if they belonged to one type. The effect that metaphor type has on reading/response times or plausibility judgements then varies along 3 degrees of freedom between the full and the null model. Hence when the full model is compared to the null model through χ^2 likelihood ratios, there are 3 degrees of freedom. Whenever we are analysing individual differences in reading/response times or plausibility judgements between two metaphor types (a pairwise comparison), there is 1 degree of freedom between the full and the null model (2 types in the full model, 1 type in the null model). Another requirement is that the two models be nested models, i.e. that the more complex model can be transformed into the other one. In our case, the full and the null model are nested models because the null model contains the same random effects, random intercepts, random slopes, and error term as the full model. Only the fixed effect of the full model, metaphor type, is absent from the null model. The null model, being the less complex one, is thus nested within the more complex full model.

Recall we defined cross-linguistic metaphor communicability in a strong and a weak sense: L2 language-specific metaphors are weakly communicative if they are more communicative than novel metaphors and they are strongly communicative if they are as communicative as cross-linguistically shared metaphors. When communicability is measured through reading/response times, faster reading/response times indicate more communicability; reading/response time delays relative to controls (the cross-linguistically shared metaphors) indicate difficulty in inferential processing and thus less communicability. Table (7.3) give the significance of differences

in reading/response times for monolinguals and bilinguals. For English monolinguals, reading/response times for German metaphors are as slow as for novel metaphors ($\chi^2(1) = .8347, p = .361$). Reading/response times for German metaphors are significantly slower than L1-specific English metaphors ($\chi^2(1) = 8.985, p = .0027$) and cross-linguistically shared metaphors ($\chi^2(1) = 19.901, p = .0000082$). Therefore, L2 language-specific German metaphors are not cross-linguistically communicative to monolinguals, neither in the strong nor the weak sense. Reading/response times for German metaphors being as slow as for novel metaphors ($\chi^2(1) = .8347, p = .361$) suggests that inferential processing is as complicated as it is for novel metaphors where speakers cannot make use of inferential shortcuts. Bilinguals, who would be familiar with the idiomatic meanings of L2 metaphors, do not show slower reading/response times for German ($\chi^2(1) = 1.1436, p = .2849$) or English metaphors ($\chi^2(1) = 2.8481, p = .0915$) than for cross-linguistically shared ones; reading/response times for German ($\chi^2(1) = .522, p = .47$) and English metaphors ($\chi^2(1) = 2.9846, p = .0841$) are also significantly faster than for novel metaphors. As expected, language-specific metaphors, to bilinguals, are as communicative as cross-linguistically shared metaphors.

In Figure (7.1) and Table (7.1), we see that median reading/response times, which are less sensitive to outliers than the mean, are around 4000 ms or less. In an effort to improve the fit of the linear mixed effects models, outliers with reading/response times > 8000 ms were excluded. Whenever we exclude outliers, we need to be wary how much observations are affected and how their exclusion affects the explanatory power of tests of statistical significance. Removing outliers from the mixed effects model fitted to reading/response times of English monolinguals resulted in $1 - 570/588 = 3.1\%$ of observations being excluded. The amounts of reading/response times excluded per metaphor type are as follows: $1 - 70/70 = 0\%$ of English metaphors, $1 - 156/168 = 7.1\%$ of German metaphors, $1 - 181/182 = 0.5\%$ of cross-linguistically shared metaphors, and $1 - 163/168 = 3\%$ of novel metaphors. Note that most of the reading/response times > 8000 ms excluded from the linear mixed effects model of English monolinguals excluded observations for German metaphors. Thus by excluding slow reading/response times as outliers, we have actually made it harder for the reading/response times of German metaphors to come out as significantly slower. But even with these slower reading/response times excluded, English monolinguals still have significantly slower reading/response times for German metaphors ($.0000082 < p < .0027$, Table 7.3).

Excluding reading/response times > 8000 ms from the model of bilinguals' responses resulted in $1 - 427/492 = 13.2\%$ of observations being removed, with reading/response times excluded per metaphor type as follows: $1 - 51/60 = 15\%$ of English metaphors, $1 - 115/144 = 20.1\%$ of German metaphors, $1 - 141/156 = 9.6\%$ of cross-linguistically shared metaphors, and $1 - 120/132 = 9.1\%$ of novel metaphors. In Figure (7.1), it seems there is a slight tendency for bilinguals to have slower reading/response times for language-specific metaphors than for non-language-specific ones. We see that by excluding reading/response times > 8000 ms as outliers, we have eliminated more reading/response times of language-specific than of non-specific metaphors. We have therefore decreased the chance for this slight tendency to be statistically significant and, indeed, the fitted linear effects model finds the difference between language-specific and non-specific metaphors to be non-significant. In Figure (A.5) in the Appendix, we see that excluding reading/response times > 8000 ms actually does give the mixed effects model a better fit to the data. It would thus seem we find ourselves in a position where we have to make a choice: we can either (1) choose to exclude fewer reading/response

times of language-specific metaphors but as a result the linear mixed effects model and tests of statistical significance would lose explanatory power or (2) we can opt for a better fit of the mixed model, which gives the tests of statistical significance more explanatory power, but by removing slower reading/response times we might remove the very trend we wish to test. I think the better option is a third, alternative one: to revisit the experimental design and implement a stronger incentive for bilinguals to respond more quickly and as quickly as monolinguals.

In Chapters 5 and 6, I proposed quantitative estimators of cross-linguistic differences in metaphors' linguistic form, familiarity of that form, and their linguistic meanings, which are summarised in Table (6.3). We might assume these differences in the material could be the source of reading/response time differences between metaphor types. Differences in familiarity are the smallest potential source because they only amount to 3.4%, whereas cross-linguistic differences in metaphors' form and meanings amount to 13.9% and 13.8% respectively, and differences in familiarity are equally distributed across metaphor types (E: 3.8%, G: 3.3%, S: 3.7%, N: 1.5% with a significance of variation between metaphor types of $p \approx .531$ ¹). When we look at the cross-linguistic differences of metaphors' form and meanings per metaphor type (E: 18.2%, G: 17.9%, S: 12%, N: 7.6% for metaphors' form² and E: 13.5%, G: 19%, S: 14.6%, N: 8.3% for their meanings) it might seem that there are slightly more differences for German-specific metaphors, which would match the result that English monolinguals have slower reading/response times for German metaphors, but the variation of differences in form and meanings is (a) non-significant ($p \approx .241$ for form differences and $p \approx .502$ for semantic differences, including relevance expectations) and (b) they are as much bigger for German than English and shared metaphors as they are for novel metaphors compared to English and shared ones. reading/response times do not seem to correlate with cross-linguistic differences in metaphors' form and meanings because if they did, novel metaphors, which have cross-linguistic differences that are even smaller than the English and shared metaphors, should have reading/response times that are even faster than those of English and shared metaphors. Instead, English monolinguals' reading/response times are as much slower for novel metaphors than English and shared metaphors as they are for German metaphors.

The German-specific metaphors, whose metaphor source and target concepts, relevance expectations, and language-specific lexical meanings are cross-linguistically approximately conceptually similar to English, should have been read as fast as English and shared metaphors. The significantly slower reading/response times for German metaphors fit the predictions made by RT and GS but not those by CMT. The fact that only monolinguals but not bilinguals had significantly slower reading/response times for L2 language-specific metaphors tells us that bilinguals must have been able to make use of inferential shortcuts that were not available to monolinguals. If metaphors' source and target concepts are cross-linguistically approximately similar between English and German, then both monolinguals and bilinguals draw inferences from similar conceptual knowledge (this seems to be true as we shall see in the results of the context creation task). If the difference in knowledge between monolinguals and bilinguals is not conceptual, then one potential explanation would be that bilinguals had knowledge of the idiomatic meanings of the L2 language-specific metaphors whereas monolinguals did not and

¹The percentages given are the averages of cross-linguistic differences in metaphors' synchronic frequency, i.e. cumulative familiarity probability, diachronic frequency change, and cumulative familiarity probability of the linguistic form of their source and target concepts taken from Table (6.3).

²The percentage given for each is the average of difference in the number of syllables, words, morphemes, and orthographic characters per metaphor type taken from Table (6.3).

this would mean that the information that is part of these idiomatic meanings which facilitated inferential shortcuts for bilinguals but not monolinguals is knowledge of non-conceptual, conventionalised linguistic meanings.

7.4.2 Plausibility judgements

Some of the English-German bilinguals were tested in the lab, the majority was tested via the internet. We therefore need to first make sure that it is appropriate to combine their results. In experiments conducted via the internet, the researcher has less control over the conditions in which participants perform the experimental tasks. These concerns go from differences in participants' hardware to their state of attention during the experiment. Comparison of the data of experiments conducted in the lab and those conducted via the internet showed that they are *qualitatively* and *quantitatively* comparable. For each metaphor type, plausibility judgements are comparable in magnitude. We can therefore conclude that, for all intents and purposes of the present investigation plausibility judgements collected via the internet are as reliable as those collected in the lab.

Metaphors are communicative if they are intelligible and in order to be intelligible they must first be plausible. Therefore, a metaphor whose implied analogy between source and target concept is implausible is also not intelligible and not communicative. The analogy is plausible only if speakers feel justified making the necessary assumptions as part of pragmatic enrichment. Metaphor plausibility judgements can therefore be viewed as an index of their intelligibility and communicativeness. Because of the closeness of language contact between English and German and the concepts and conceptual metaphors that are shared between them because of this closeness, when speakers of both languages share the basic concepts and conceptual metaphors for a particular metaphoric proverb, that proverb should be as intelligible to non-native speakers as it is to native speakers and this, CMT would predict, should be true by virtue of conceptual plausibility. RT and GS, however, would predict that the proverb should not be cross-linguistically intelligible on the basis of conception alone, but only if speakers know the linguistic conventions associated with it: the idiomatic meaning, omitted context, and expectations of contextual relevance. RT and GS would thus predict that non-native speakers should not find proverbs from the other language as plausible as proverbs from their native language because their considerations of 'plausibility' include their knowledge of linguistic conventions (or lack thereof) *on top of* purely conceptual considerations and they would predict that considerations of linguistic convention should take precedence over purely conceptual considerations.

In the out-of-context experiments, participants gave ratings on 5-point Likert scales; in the in-context experiments, on 7-point scales. We can thus normalise 5-point ratings and their standard deviation by subtracting the minimal rating possible, 1, and dividing by 4 and 7-point ratings and standard deviation by subtracting 1 and dividing by 6. The graduation of the scale was changed in hopes to get more fine-grained ratings, which ultimately turned out to make not much of a difference. Nevertheless, conversion of the original plausibility ratings to z-scores allows for direct cross-experimental comparison. We can also normalise ratings by converting them to z-scores

$$z = \frac{x - \mu}{\sigma}, \quad (7.5)$$

where μ is an experiment's sample mean, σ is an experiment's sample standard deviation in responses, and x is any measured response that we wish to convert to a z-score. In our case,

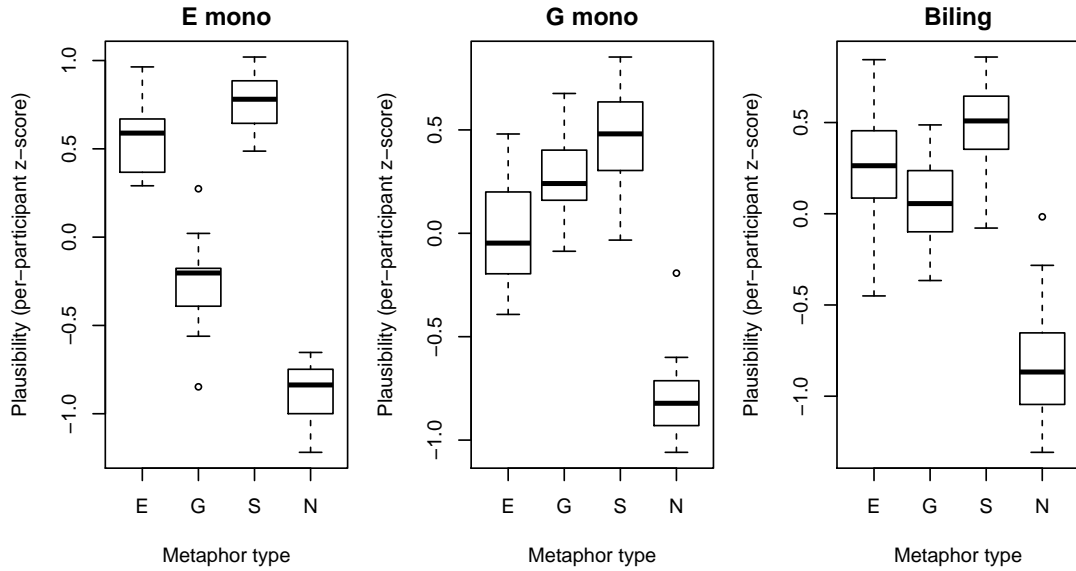


Figure 7.3: The y-axis shows median plausibility ratings per metaphor type, where ratings are converted to per-participant z-scores. Boxes and whiskers indicate between-participant variation in z-scores.

z-scores are comparable across experiments because the sample mean and standard error in ratings give similar estimates of the likely population parameters. The experimental sample means are similar across experiments: out-of-context plausibility judgements, E mono: .542, G mono: .526, Biling: .525; in-context plausibility judgements, E mono: .584, G mono: .579, Biling: .532. Experiments with more participants should give us estimates of the sample mean that are closer to the true population mean than experiments with fewer participants. The average experimental sample mean $\bar{\mu}$, weighted by the number of participants w_i , is

$$\bar{\mu} = \frac{\sum w_i \mu_i}{\sum w_i} \approx .551.$$

One standard deviation in ratings, where 95% of all ratings lie, is also similar across experiments: out-of-context experiments, E mono: .342, G mono: .332, Biling: .338; in-context experiments, E mono: .394, G mono: .365, Biling: .375. The average sample deviation in ratings $\bar{\sigma}$, weighted by the number of participants w_i , is

$$\bar{\sigma} = \frac{\sum w_i \sigma_i}{\sum w_i} \approx .362.$$

Given in Figure (7.3) are the median plausibility ratings per metaphor type as per-participant z-scores. The variation indicated in the graphs in Figure (7.3) is the variation in the per-metaphor-type z-scores *between* participants.

In the analysis of the reading/response times, the main reason for using linear mixed effects models instead of ANOVAs was that a per-participant ANOVA ignores by-item variation (by averaging across multiple items of a particular metaphor type) and a per-item ANOVA ignores by-subject variation (by averaging across multiple responses of a metaphor item by different

participants). Linear fixed-effects-only models and ANOVAs do not allow us to consider both by-subject and by-item variation in the same analysis; they therefore falsely treat responses per metaphor type by multiple participants as independent of each other and falsely treat multiple responses by the same participant as independent of each other. The same reasoning straightforwardly applies to plausibility judgements as well. Some participants might generally refrain from ever using the lowest ratings, while other participants might not, and yet others might generally avoid giving the highest ratings. As a result, the plausibility judgements, regardless of metaphor type, are not independent of each other but depend on a participant’s baseline rating. At the same time, there are multiple metaphors per metaphor type and so if metaphor type actually does have an effect on plausibility judgements, we should expect that metaphors of the same type should receive similar plausibility ratings, but of course by-item variation in plausibility judgements would also depend on by-subject variation in baseline ratings. Consequently an analysis that adequately accounts for these two non-independences of data points must consider by-subject and by-item variation, their interaction with metaphor type, and do so without averaging across either one. An analysis using linear mixed effects models is able to do that, but linear fixed-effects-only models and ANOVA are not. Plausibility judgements in this and the next chapter are therefore analysed by fitting linear mixed effects models to the judgements of English and German monolinguals and English-German bilinguals using the following model formula:

$$y = X\beta + Z_1b_1 + Z_2b_2 + \varepsilon \quad (7.6)$$

where y is a vector of observations, the *plausibility judgements*, β is the fixed effect, *metaphor type*, b_1 is by-subject variation in plausibility judgements y (as a random effect) and b_2 is by-item variation in plausibility judgements y (also as a random effect). X , Z_1 , and Z_2 are model matrices of regressors relating the observed plausibility judgements y to the fixed effect β and the random effects b_1 and b_2 . Matrix X contains random intercepts; Z_1 and Z_2 contain random intercepts *and* random slopes, i.e. the model expects different baseline plausibility ratings (the intercepts) and magnitudes of the effect of metaphor type on these ratings (the slopes) *per subject* and *per metaphor item* within each metaphor type.

Visual inspection of the residual plots did not reveal any obvious deviations from homoscedasticity (Figure A.6) or normality (Figure A.7). Visual inspection also showed no evidence of non-linearity, ceiling or floor effects in the data. Using *linear* mixed effects models on the untransformed Likert scale ratings should therefore be appropriate. Shown in Figure (A.6) are the residual plots for the linear mixed effects model fitted to the out-of-context and in-context plausibility judgements. The stripes are a result of the Likert scale used to record plausibility judgements: a 5-point Likert scale for the out-of-context condition and a 7-point scale for the in-context condition. However, notice that each stripe spans all fitted values and at least half the residuals. This means that although plausibility ratings are restricted to the increments of the respective Likert scale, within that increment, by-subject and by-item variation *is* homoscedastic. Figure (A.7) shows the probability densities of residuals, i.e. the variation of plausibility judgements about the regression lines of the fitted mixed effects models for English monolinguals, German monolinguals and English-German bilinguals. The histograms in the left column are the residuals of the out-of-context plausibility judgements; in the right column are the residuals of the in-context plausibility judgements. The actual kernel density of plausi-

Table 7.4: Variance explained by each fixed and random effect in the linear mixed effects models: out-of-context plausibility judgement models. Each value is a ratio of the variance accounted for by that effect and the total model-wide variance. These ratios are effectively R^2 estimates ranging from 0 to 1 (for ease of comparison, given here as percentages). The higher the value, the more of the overall variance is explained by the respective effect.

Effect	E mono	G mono	Biling
Metaphor type	68.3%	47.5%	50.0%
Subject	12.5%	4.1%	8.6%
Item	10.8%	26.9%	22.4%
Error term	8.4%	21.5%	19.0%

bility judgements is plotted as a continuous line. The corresponding normal distribution, i.e. a normal distribution using the mean and standard deviation of the actual distribution, is plotted as a dashed line. The actual residuals (as given in Figure A.6) are plotted as a rug along the x-axis. By comparing the kernel densities to the normal densities we can see that monolinguals' and bilinguals' plausibility judgements are normally distributed. The normality requirement for their linear mixed effects models is therefore satisfied. The linearity requirement is met and because there is only one fixed effect, non-collinearity is also satisfied.

The vector of random error terms ε in Formula (7.6) is the proportion of variance in plausibility judgements y that is not accounted for by either the fixed or random effects. Since only the fixed effect, metaphor type, and the random effects, by-subject and by-item variation, are explicitly modelled, ε is implicitly given by the residual of the fitted model, i.e. the variance in plausibility judgements y that is left “unexplained” by the fitted model. Thus the proportion of the variance in plausibility judgements y explained by the fixed and random effects and the “unexplained” variance ε tells us how well a linear mixed effects model fits observation y . The mixed effects model fitted to the judgements of English monolinguals accounts for 91.6% (see Table 7.4, 68.3% + 12.5% + 10.8%) of the variation in metaphors' plausibility judgements. The model fitted to the ratings of German monolinguals accounts for 78.5% (47.5% + 4.1% + 26.9%) of the total variation in plausibility judgements. The model fitted to the judgements of English-German bilinguals can explain 81% (50.0% + 8.6% + 22.4%) of the variation in plausibility judgements. For all speaker groups, the effect we are interested in, *metaphor type*, accounts for more of the total variance in plausibility judgements than either by-subject (68.3% > 12.5%, 47.5% > 4.1%, 50.0% > 8.6%) or by-item variation (68.3% > 10.8%, 47.5% > 26.9%, 50.0% > 22.4%) and more than by-subject and by-item variation taken together (68.3% > 12.5% + 10.8%, 47.5% > 4.1% + 26.9%, 50.0% > 8.6% + 22.4%). Hence, most of the variation in plausibility judgements that the models fit so well is due to the fixed effect *metaphor type* and not due to the random effects.

$$y_0 = Z_1 b_1 + Z_2 b_2 + \varepsilon \tag{7.7}$$

The null model, shown in Formula (7.7), is derived from, and therefore nested within, the full model. It is the full model *without* the fixed effect, metaphor type. It otherwise contains the same random effects, random intercepts, and random slopes as the full model. I follow the

Table 7.5: Out-of-context plausibility judgements on 5-point Likert scale. Best-fitting regression using mixed-effect modelling. Models fitted to responses of English monolingual native speakers (E mono, $R^2 = .916$), German monolingual native speakers (G mono, $R^2 = .785$), and English-German bilinguals (Biling, $R^2 = .810$). Intercept for metaphor type is cross-linguistically shared metaphors. P-values are likelihood ratios (χ^2 tests) between the linear mixed effects model that includes the variable of interest, *metaphor type*, and the model that does not (the null model). Significance codes: *** for $p \leq .001$, ** for $p \leq .01$, * for $p \leq .05$, † for $p \leq .1$.

Speaker	Predictor	B	St err	t	p
E mono	S (intercept)	4.3432	.1610	26.976	< .0001 ***
	E (L1)	-0.2509	.2322	-1.081	.27
	G (L2)	-1.6124	.3439	-4.689	.000057 ***
	N	-2.5040	.2514	-9.962	< .001 ***
G mono	S (intercept)	3.7795	.2337	16.173	< .0001 ***
	E (L2)	-0.7405	.3850	-1.923	.022 *
	G (L1)	-0.2457	.2770	-0.887	~1
	N	-1.9518	.3220	-6.061	< .001 ***
Biling	S (intercept)	3.8521	.1928	19.981	< .0001 ***
	E (L2)	-0.3670	.3119	-1.177	.303
	G (L1)	-0.6507	.2603	-2.500	.026 *
	N	-2.0294	.2829	-7.173	< .001 ***

recommendation of Schielzeth & Forstmeier (2009) and Barr et al. (2013) here, who argue that such models are less prone to Type I errors. Given Wilks Theorem (Wilks, 1938), the likelihood ratio between two nested models, the full and the null model, approaches a χ^2 distribution, which can thus be used to calculate p-values. Note that in our particular case, because the full model has only one fixed effect, metaphor type, and the null model has only random effects, the null model is our way of sampling the random distribution of metaphor plausibility judgements when metaphor type is assumed to play no role. Comparing the full model, which is “aware” of a metaphor’s type, and the null model, which is “unaware” of a metaphor’s type, is thus equivalent to asking whether the plausibility judgements that are given with an “awareness” of metaphor type are normally (i.e. randomly) distributed. Hence, in our particular case, the p-values we get from the χ^2 likelihood ratio tests are notionally directly comparable to classical ANOVA, only that, unlike per-participant and per-item ANOVAs, our mixed effects models do not violate the assumed independence of data points. These are the p-values given in Table (7.5).

We are interested in the cross-linguistic communicativeness of L2 language-specific metaphors. Recall that we defined metaphors’ cross-linguistic communicativeness in two ways: L2-specific metaphors are strongly communicative if they are judged to be as plausible as L1-specific and cross-linguistically shared metaphors. They are weakly communicative if they are judged to be more plausible than unfamiliar novel metaphors. If CMT is correct in its predictions, then English and German monolinguals should find L2-specific metaphoric proverbs strongly communicative. However, as we can see from Table (7.5), English monolinguals judge L2 language-specific German metaphors to be significantly less plausible than L1-specific English metaphors ($\chi^2(1) = 11.060, p = .000882$) and significantly less plausible than cross-linguistically

shared metaphors ($\chi^2(1) = 16.211, p = .000057$). We know comparing L2-specific German metaphors to L1-specific metaphors is an adequate measure of L2 metaphors' cross-linguistic communicability because English monolinguals found L1-specific English metaphors as plausible as cross-linguistically shared metaphors ($\chi^2(1) = 1.219, p = .270$). If CMT is correct, then German monolingual native speakers should find L2 English-specific metaphors as plausible as L1-specific German metaphors and cross-linguistically shared ones. However, as we saw with English monolinguals, monolingual German native speakers give significantly lower plausibility ratings to L2-specific English metaphors than they give to L1-specific German metaphors ($\chi^2(1) = 8.286, p = .00399$) or cross-linguistically shared ones ($\chi^2(1) = 5.244, p = .022$).

We know comparing L2-specific English metaphors to L1-specific German ones is a measure of strong cross-linguistic communicability that is comparable to comparing the plausibility ratings of English and cross-linguistically shared metaphors because L1-specific German metaphors and cross-linguistically shared ones received similar plausibility ratings ($\chi^2(1) = 0, p = 1$). Contrary to what CMT predicted, even when the source and target concepts of L2 language-specific metaphors, their lexical meanings, their relevance expectations, the salience and familiarity of their linguistic form, and related metaphoric imagery (e.g. *the pen is mightier than the sword* uses a metaphoric image similar to the language-specific English proverb *word are daggers* that was tested) are cross-linguistically similar, monolingual native speakers do not find L2 language-specific metaphors as plausible as L1 and cross-linguistically shared metaphors they know from their native language. Monolinguals find L2-specific metaphors only cross-linguistically communicative in the weak sense. English monolinguals find L2-specific German metaphors to be significantly more plausible than unfamiliar novel metaphors ($\chi^2(1) = 7.138, p = .00755$). German monolinguals find L2-specific English metaphors significantly more plausible than novel metaphors ($\chi^2(1) = 9.468, p = .00209$).

Notice that the analysis is biased towards confirming CMT because in order for L2 language-specific metaphors to be classified as cross-linguistically communicative we do not require that speakers actually infer the same meanings as native speakers, all we require is that whatever meanings they do infer, these meanings be plausible. Therefore, it was even more likely that statistical analysis should find that monolinguals rate L2-specific metaphors to be as plausible as L1-specific and cross-linguistically shared ones. However, the results strongly contradict Conceptual Metaphor Theory and instead support Relevance Theory and Graded Salience.

Bilinguals serve as a control group. If, as CMT claims, plausibility intuitions and intuitions of meaning salience give directly insight into the make-up of speakers' conceptual systems because salient meanings are in fact primary embodied concepts intertwined through *conceptual metaphors* into ever more complex higher-order inferential structures, then bilinguals' plausibility judgements should support this. As we shall see in the context creation task, and as we already expect from the corpus analysis of Chapter 5, bilinguals know the idiomatic meanings of all of the language-specific metaphors. So according to CMT, if plausibility judgements directly reflect meaning salience and conceptual structure, then bilinguals should give plausibility ratings to language-specific metaphors that are as high as those of cross-linguistically shared metaphors because we know the source and target metaphor concepts of language-specific proverbs are as cross-linguistically similar as those of the cross-linguistically shared proverbs (see Table 6.3, cross-linguistic approximate conceptual similarity of English-specific metaphors is 86.5%, 81% for German-specific metaphors, and 85.4% for cross-linguistically shared metaphors, and the variation between metaphor types is non-significant, $p \approx .502$) and we know bilinguals are

familiar with their idiomatic meanings. However, experimental results show that bilinguals only find English language-specific metaphors as plausible as cross-linguistically shared ones ($\chi^2(1) = 1.061, p = .303$), but not German-specific metaphors ($\chi^2(1) = 4.948, p = .026$). Although lower, the plausibility ratings of German-specific metaphors are not as low as those of novel metaphors ($\chi^2(1) = 15.403, p = .000087$). In the next chapter, we will see that there is more to this because context affects monolinguals and bilinguals differently in their plausibility judgements. For now, the finding that bilinguals find L1-specific German metaphors to be less plausible than cross-linguistically shared ones should lead us to suspect that bilinguals are expressing more than “conceptual plausibility” through their judgements. Recall, bilinguals are predominantly advanced second language learners of English and they saw all metaphors in German, same as the German monolinguals. Hence lowered plausibility of L1 metaphors cannot be the result of “bad” translation because the L1 metaphors were not translated. Bilinguals saw the L2-specific English metaphors in German translation. If translation was the cause of lower perceived plausibility, bilinguals’ ratings of English-specific metaphors should have been lowered. Also, German monolinguals rated German-specific metaphors to be as plausible as cross-linguistically shared ones ($\chi^2(1) = 0, p = 1$) and from the corpus analysis we know that the German-specific metaphors are as frequent (.11 tokens per million words, see Table 5.2) as the English-specific ones (.11 to .24 tokens per million words in American and British English, see Table 5.2). And in post-experimental interviews bilinguals were as familiar with the German-specific metaphors as German monolinguals. We therefore know that it cannot be the case that the German-specific metaphors received lower plausibility ratings because bilinguals were less familiar with them than monolinguals or because the metaphors themselves were more obscure than the English-specific proverbs. As we shall see in the next chapter, one potential explanation might be that bilinguals give lower plausibility ratings to German metaphors to express their sensitivity for metaphors’ language-specificity and the more constrained expectations of contextual relevance that come with it. In other words, bilinguals are not judging “conceptual” plausibility but metaphor plausibility in light of their knowledge of linguistic convention. Maybe bilinguals’ awareness is heightened for L2 metaphors as a result of explicit instruction and lowered for L1 metaphors because those L1 metaphors would not have been part of that instruction. L2 learning usually focuses on L2-to-L1 and not on L1-to-L2 when it comes to proverbs, sayings, and idioms, and therefore overlooks L1 metaphors, i.e. usually there is no focus on those German metaphors that have no English equivalent; the focus is on learning English, not on learning German. For that reason, bilinguals might self-consciously correct their L2 metaphor intuitions because they think they ought to be more meaningful (hyper-grammaticality, the language perception equivalent to hyper-correctness in language production) whereas they do not correct their intuitions for L1 metaphors. So really we should expect L2 metaphors to have lower ratings—and we should expect them to be as low as those of the L1 metaphors. Bilinguals express their awareness for the language-specificity of the L1 and L2 metaphors in their lower ratings, lower than shared metaphors; i.e. they make the following distinction between metaphors in the experiments.

Since monolinguals are unaware of the language-specificity of the L1 and L2 metaphors, they only distinguish between metaphors familiar and metaphors unfamiliar to them. We would have expected bilinguals to be even more likely than monolinguals to give judgements of conceptual plausibility that confirm and support CMT because they are more likely than the monolinguals to have the conceptual mappings necessary to comprehend metaphors from both English and

German and find them intelligible. However, bilinguals' judgements of conceptual plausibility do not support CMT. Moreover, their judgements and reports in interviews show an awareness for the language-specificity of English and German metaphors in contrast to metaphors that are commonly found in both languages and this awareness can thus not be interpreted as a direct spellout of bilinguals' conceptual system, but must incorporate language-specific *linguistic* knowledge. We might therefore take this as evidence that linguistic knowledge takes precedence over purely conceptual knowledge.

In summary, there are two findings in the plausibility judgements that falsify the claims made by CMT: (1) monolinguals find L2 language-specific metaphors significantly less plausible than metaphors they know from their native language (L1-specific and cross-linguistically shared metaphors) although cross-linguistic approximate conceptual similarity is maximised, and (2) bilinguals give lower plausibility ratings to L1 metaphors although they are as familiar with them as German monolinguals, although they saw the L1 metaphors in their original and conventional German phrasing, and although the German-specific metaphors are as frequent as English-specific metaphors which bilinguals rated to be as plausible as cross-linguistically shared metaphors ($\chi^2(1) = 1.061, p = .303$).

7.4.3 Metaphor frequency as covariate

When metaphors' corpus frequency is included as a covariate in the mixed effects models of English monolinguals' plausibility judgements, the correlation between metaphor frequency and metaphor type is $r = -0.483$ for cross-linguistically shared metaphors, $r = 0.363$ for English metaphors, $r = 0.185$ for German metaphors, and $r = 0.343$ for novel metaphors. Cohen (1992) provides guidelines to evaluate the strength of correlation. According to Cohen, the correlation between corpus frequency and English monolinguals' plausibility judgements of German metaphors is small, and the interaction between frequency and metaphor type is not significant ($\chi^2(2) = 3.7567, p = .153$). So metaphor frequency is not a good explanation why English monolinguals give lower plausibility ratings to German metaphors. For the other metaphor types, the correlations are medium in strength.

When metaphor frequency is included as a covariate in the mixed effects models of German monolinguals, the strength in correlation for cross-linguistically shared metaphors is $r = -0.620$, $r = 0.291$ for English metaphors, $r = 0.396$ for German metaphors, and $r = 0.521$ for novel metaphors. The correlation between metaphor frequency and German monolinguals' plausibility judgements of English metaphors is small according to Cohen (1992), smaller than for any other metaphor type, and the interaction between frequency and metaphor type is not significant ($\chi^2(2) = .1396, p = .933$). Therefore, as with English monolinguals, metaphor frequency is not a good explanation why German monolinguals give lower plausibility ratings to English metaphors.

For bilinguals, $r = -0.641$ for cross-linguistically shared metaphors, $r = 0.298$ for English metaphors, $r = 0.353$ for German metaphors, and $r = 0.519$ for novel metaphors. The correlation between corpus frequency and bilinguals' plausibility judgements of language-specific English and German metaphors is medium to small according to Cohen (1992) and smaller than for the other metaphor types. The interaction between frequency and metaphor type is not significant ($\chi^2(2) = 1.2698, p = .530$).

7.4.4 The context creation task

If comprehending a particular idiomatic metaphor requires knowledge of its idiomatic meaning and the idiomatic meaning can be thought of as the context omitted from overt linguistic expression that would be necessary to ‘spell out’ the full conventional speaker meaning of the metaphoric proverb in a non-idiomatic but ‘literal’ way, then native speakers who know the idiomatic meaning and the intended full conventional meaning should be able to make (at least part of) the omitted context explicit. Since the idiomatic meaning also includes expectations of contextual relevance, when speakers are asked to create sensible context for a metaphoric proverb, only native speakers but not non-native speakers should be able to meet the proverb’s conventional relevance expectations. If it is appropriate to think of idiomatic meanings as language-specific conceptual metaphors, then non-native speakers should be able to make relevant context explicit on the basis of their conceptual knowledge and considerations of conceptual plausibility whenever they share the basic concepts and conceptual metaphors necessary for a particular idiomatic metaphor with native speakers. CMT would propose that idiomatic meanings might be language-specific conceptual metaphors and would therefore predict that non-native speakers should be able to do this for two languages where, for the proverbs in question, basic concepts and conceptual metaphors are cross-linguistically shared because of intense language contact. RT and GS, on the other hand, see the idiomatic meaning, the knowledge of omitted context, and expectations of contextual relevance as part of arbitrary linguistic convention and not conception, and not necessarily as motivated by Embodied Cognition. RT and GS would therefore predict that non-native speakers should fail to make context explicit which is relevant to the conventional meaning of an L2 proverb.

In part 2 of the experiments, participants were asked to come up with continuations in the form of prepositional phrases to eight of the metaphors out of those they had already seen in part 1: of those eight metaphors, two were English-specific metaphors, two German-specific, two cross-linguistically shared, and two novel metaphors. It goes without saying that the interpretation of participants’ data in the context creation task is highly subjective. I therefore ask that you take the analysis that follows with the proverbial grain of salt.

In interviews conducted after the experiments, participants generally reported that they found it harder to come up with continuations for metaphors they were unfamiliar with and/or that “made no sense” to them. Monolinguals said they found it harder to create continuations for the German and the novel metaphors than for the English and the shared metaphors. Bilinguals reported they found it more difficult to come up with continuations for novel metaphors than for English, German, and shared metaphors. For those tested in the lab, I can confirm that it also took participants considerably longer to come up with continuations for the metaphors they reported posed more difficulty. A few participants specifically pointed out that they found it hard to think of something, as they put it, *creative* or *unconventional* for the metaphor they were familiar with. I assured them originality was not a requirement, but many felt obliged anyway. Nevertheless, when we look at the phrasing of the continuation we find a surprising agreement among those conventional metaphors that were familiar to participants: the English and shared metaphors for English monolinguals, the German and shared metaphors for German monolinguals, and the English, German, and shared metaphors for bilinguals. For example, for the metaphor *silence is golden*, almost all participants created some paraphrase of *speech is silver* as continuation which is the conventional second half of the proverb.

What participants did not know is that I intended to use these continuations as the context that would be given to the participants in the in-context study described in Chapter 8. This way, native speakers would be helping each other to make cross-linguistically unintelligible metaphors intelligible. Of course, since participants only created continuations for eight of the 41 metaphors, I had to create continuations for the other 42 metaphors used in the in-context study (recall from Table 5.1 that there are 50 metaphors in the in-context study), all of which were created such as to improve metaphors' intelligibility by either alluding to the conceptual metaphoric mapping or by highlighting features of the source and target concept that should be transferred from one to the other—but I shall go more into detail about this in Chapter 8 and leave you with this outlook.

Continuations created by English monolinguals

Let's take a look at the continuations created for cross-linguistically shared, language-specific, and novel metaphors in turn.

Cross-linguistically shared metaphors

- (9a) Ignorance is bliss for the individual / for some / to many. (**generic**)
- (9b) Ignorance is bliss for the mind / when the truth is troubling. (**state of mind**)
- (9c) Ignorance is bliss for a short while / to a child. (**of limited duration**)
- (9d) Ignorance is bliss for the ignorant one / to the ignorant / for the unknowing / to the naïve / to a child. (**naivety**)

I have listed *to a child* twice, under two different senses, because it could be interpreted in two ways: (1) a child as being more naive, ignorant, and uninformed about the world compared to an adult, or (2) in the sense that childhood is a limited period in one's life and that as we grow up the solace offered by ignorance as the innocence of youth is lost or outgrown. The senses are ordered according to their frequency, from the first one being the most frequent and then decreasing. Hence, the generic and state of mind sense occurred most often.

- (10a) Silence is golden in a library / in a solemn situation. (**institutional**)
- (10b) Silence is golden after an hour with the kids / to parents with loud children / for the tired / in a noisy room / when you need peace. (**interpersonally and personal needs**)
- (10c) Silence is golden if you have nothing nice to say / after a fight. (**not advisable**)
- (10d) Silence is golden in moderation / in the morning / to those who want it. (**preferences vary**)

The first sense is the one that is most common among the continuations English monolinguals created. Note that senses alternate between two opposing views of silence: (1) silence as something *positive*, something that provides peace of mind, allows one to think, work, be productive, keeping silent as a means to protect oneself or others, or (2) silence as something *negative*, silence as a sign of displeasure, discomfort, disgust, or resentment, or as something that is only appropriate in moderation or in certain situations. And note that they contrast in

terms of (1) keeping silent in order to obey a social norm or (2) keeping silent as a means to an end (e.g. *if you have nothing nice to say*).

(11a) Anger is a volcano inside a person / for emotions / in your emotions / of an emotion / of emotion / of hatred / of passion. (**emotionally**)

(11b) Anger is a volcano on the verge of eruption / waiting to erupt / which could explode at any moment. (**on the verge of eruption, holding back anger**)

(11c) Anger is a volcano no longer dormant / in action. (**erupting, venting anger**)

The emotional sense was by far the one that continuations alluded to most often. Note that the other two senses are entailed by the general emotional sense, but in such a way that they focus on the salient features that anger and volcano shared, i.e. the features that constitute the metaphoric analogy between source and target concept.

English-specific metaphors

(12a) Love is a journey in life / on the road of life. (**a part of life**)

(12b) Love is a journey through good times and bad / with ups and downs / which is long and harsh. (**ups and downs**)

(12c) Love is a journey of the heart / to(wards) happiness. (**emotionally**)

(12d) Love is a journey to marriage / through good times and bad / with ups and downs. (**marriage, wedding vows**)

(12e) Love is a journey with someone. (**interpersonally**)

Love is a journey in life is the sense that continuations alluded to most often. Recall that in the previous chapter I argued that native speakers would see love as a part of life and would therefore conceptually make a connection between *life is a journey* and *love is a journey*. We see here that English monolinguals seem to have made that connection because they give *in life* as a continuation for *love is a journey*. Note that I have listed *through good times and bad* twice because (1) it is part of the *love is a part of life* analogy and (2) it is part of the conventional phrasing of wedding vows.

German-specific metaphors

(13a) Sports is murder after 20 mins / for the weak / for the body / on your legs. (**straining, exhausting**)

(13b) Sports is murder when you're bad at it. (**the non-athletic**)

(13c) Sports is murder to fair competition / competitive and violent. (**loosing**)

(13d) Sports is murder for some animals. (**hunting**)

(13e) Sports is murder in some countries. (**literal**)

Most English monolinguals gave a continuation suggesting that sports is straining or exhausting or took *murder* as a euphemism for loosing in a competitive sport. The only English speakers whose continuation in (13b) came closest to the German idiomatic meaning was a bilingual (English L1). The *straining* and *exhausting* sense is part of the German idiomatic meaning, but as I said in the previous chapter, it is not surprising that English speakers would conceptually come to this sense. However, without the additional assumption that it is straining and exhausting because a person is non-athletic is key. Without it, it is not the idiomatic meaning that German speakers would have. We see at least three senses that come nowhere near the German idiomatic meaning: taking *murder* as a euphemism for loosing a match seems conceptually plausible, one might even say it would be a consequence of the idiomatic meaning, but it is certainly not part of it. The continuation *in some countries* suggests that the participant took *murder* literally and the same seems to apply to *for some animals* where the mental image is more that of game hunting as a sport.

- (14a) The devil is a squirrel if you own nut trees / to a harvester of nuts / to nuts / to an acorn / with too many acorns / on acid. (**squirrels as nuisance**)
- (14b) The devil is a squirrel in disguise / with horns / with a stash of nutty temptation. (**the devil disguised as a squirrel**)
- (14c) The devil is a squirrel in squirrel mythology / a small meaningless animal / on the washing-line. (**confused**)

The devil is a squirrel certainly was the metaphor that confused English participants the most. The overwhelming majority seem to take *devil* as a euphemism for *nuisance*: *because it is annoying, to our conscience*. Others seem to only focus on traits of squirrels that make them a nuisance: *if you own nut trees, to a harvester of nuts, to nuts, to an acorn, with too many acorns, on acid*. Recall that the German idiomatic meaning takes the squirrel as an omen of sudden misfortune. I would argue that in the continuation that talk about the squirrel as a nuisance it is not an omen and thus they are not similar to the German idiomatic meaning. Other continuations are clearly signs of confusion where participants did not know what to make of the metaphor at all: *in squirrel mythology, on the washing-line, a small meaningless animal?* (the question mark is part of the continuation the participant gave), or the rather cruel *in a blender*. The only three that at least try to take the implied metaphoric analogy serious are quite general characterisations: *in disguise* comes closest to the squirrel being an omen, but arguably *disguise* here does not mean *omen*. The continuation *with horns* clearly picks out a conventional characteristic of the devil, but similarly as the one before seems to suggest that the participant was imagining the devil disguised as a squirrel. The third, *with a stash of nutty temptation*, although it picks out a prominent conventional feature of the devil, temptation, this mental image seems quite harmless as far as temptations go, e.g. compared to carnal temptations that would be much stronger.

Novel metaphors

- (15a) A home is money for investment / for the future / in a different form / in the bank / on the property market / to an estate agent. (**an investment in the future**)
- (15b) A home is money for the family / for the soul. (**personally**)

(15c) A home is money absorbing / down the drain. (**costly**)

In most cases the continuations that English monolinguals created alluded to a home as an investment in one's future, closely followed by continuations that alluded to personal benefits. Continuations that also considered the price of that investment as costly occurred rarely, but did occur none the less. When we compare this to the cross-linguistically shared metaphors of *silence is golden* and *ignorance is bliss*, we see that for the novel metaphor *a home is money*, speakers also contrast positive and negative interpretations of the metaphoric meaning and for both types of senses, the features that they focus on are the conceptually most salient features that successfully map from the metaphor source concept *money* onto the target concept *home*.

(16a) A friend is a magnet for fun / for happiness / for good times / in times of trouble.
(**emotionally**)

(16b) A friend is a magnet of honesty / of trust / to keep you on the right path. (**morally**)

(16c) A friend is a magnet for other friends. (**interpersonally**)

So far for the other metaphors there was always a group of continuations that alluded to a particular sense and that group occurred much more often than continuations alluding to other senses. However, in the case of *a friend is a magnet*, continuations alluding to the three senses were quite evenly distributed. That continuations would allude to the emotional and interpersonal sense is not necessarily surprising. That continuations would allude to a moral sense, however, might be surprising, after all, what does a magnet have to do with morality? Recall that I pointed out in the previous chapter that the concept of friendship ties into morality because friendship is a specific species of interpersonal and mutual moral obligations. Hence if friendship has a moral sense as part of its lexical meaning then this makes a salient candidate for the metaphoric analogy, but only if it can successfully pick out matching features of the concept *magnet*. Also note that continuations of the other two senses also link back to the emotional and interpersonal senses that are part of the lexical meaning of *friend*. The continuations that see *a friend is a magnet* in a moral sense, construct the metaphoric analogy as follows: a friend who has a lot of influence on building one's character, in a moral sense, attracts those moral qualities and virtues that are desirable (e.g. *honesty, trust*) in the same way that a magnet attracts ferromagnetic metals, which might also be desirable. Hence, while the continuations certainly relate to the lexical senses of *friend* and *magnet* individually, they also relate to the implied metaphoric analogy of the original metaphor by extending that analogy, that a friend attracts things similarly to how a magnet attracts things, in a conceptually salient way.

Continuations created by German monolinguals

In analogy to the order in which I discussed the continuations created by English speakers, we will look at cross-linguistically shared metaphors first, then language-specific, and finally novel ones.

Cross-linguistically shared metaphors

(17a) *Unwissenheit ist ein Segen von ungestörtem Glück* / *wenn das*
Ignorance is a blessing of undisturbed luck/happiness / when the
Wissen wehtun würde.
knowledge hurt would.

‘Ignorance is bliss of undisturbed happiness / when the knowledge would hurt.’
(**generic**)

(17b) *Unwissenheit ist ein Segen für den Seelenfrieden.*
Ignorance is a blessing for the piece of mind.
‘Ignorance is bliss for one’s piece of mind.’ (**state of mind**)

(17c) *Unwissenheit ist ein Segen für Dumme / für den Dummen / für den Unwissenden.*
Ignorance is a blessing for dumb-ones / for the dumb / for the unknowing.
‘Ignorance is bliss for ignorant ones / for the ignorant / for the unknowing.’ (**naivety**)

(17d) *Unwissenheit ist ein Segen auf kurze Zeit / auf den man sich nicht verlassen sollte.*
Ignorance is a blessing on short time / on which one oneself not rely-on should.
‘Ignorance is bliss for a short time / which one should not rely on.’ (**of limited duration**)

(17e) *Unwissenheit ist ein Segen mit Tücken / mit einem Fluch / nur für die Unwissenden.*
Ignorance is a blessing with perils / with a curse / only for the unknowing.
‘Ignorance is treacherous bliss / cursed bliss / bliss (but) only for the unknowing.’ (**false solace**)

The generic, state-of-mind, naivety, and of-limited-duration sense are the four senses that match the ones that English monolinguals came up with. Hence, for these four senses, English and German monolinguals associate the same senses with the meaning of the metaphor. However, a few German monolinguals also created continuations that alluded to a sense that English monolinguals did not allude to: the false-solace sense. Note that *Seelenfrieden* is also used in a non-religious sense, so although *Seele* literally means *soul*, in most context it is the modern German equivalent of English *piece of mind*.

(18a) *Schweigen ist Gold in den Momenten der Andacht.*
Silence is gold in the moments of the solemnity.
‘Silence is golden in a solemn situation.’ (**institutional**)

(18b) *Schweigen ist Gold und Reden ist Silber.*
Silence is gold and speech is silver.
‘Silence is golden and speech is silver.’ (**conventional**)

(18c) *Schweigen ist Gold im Lärm der Worte.*
Silence is gold in the noise of the words.
‘Silence is golden amidst the noise of words.’ (**interpersonally and personal needs**)

(18d) *Schweigen ist Gold für schlaue Leute / für die Herrschenden / mit dem man sich viel erkaufen kann.*
Silence is gold for smart people / for the reigning ones / with which one oneself much buy can.
‘Silence is golden for smart people / for leaders / with which to advance one’s station in life.’ (**withholding knowledge is power**)

- (18e) *Schweigen ist Gold ohne Kredit / von dem man nicht reicher wird.*
 Silence is gold without credit (loan) / from which one not richer becomes.
 ‘Silence is golden without credit or loan / that does not make you richer.’ (**no actual gold**)

Note that in (18d), we have *erkaufen* (*to buy*) with the prefix *er-*, not *kaufen* (*to buy*). The form with the prefix exclusively refers to *buying* in a non-economic but metaphorical sense as in English *to buy us some time* (*sich Zeit erkaufen*) or where it implies bribery as in *he bought his vote* (*er hat sich seine Stimme erkauft*). Hence the continuation *mit dem man sich viel erkaufen kann* cannot refer to the purchase of actual goods for money, but means advancing one’s station in life. Recall that English monolinguals also alluded to the sense that personal preferences as to when it is prudent to keep silent may vary. The German monolinguals did not focus on this sense, but instead created continuations for two related senses: (1) that withholding knowledge by keeping one’s silence can be power, too, and (2) they emphasised that actual gold or money is not at the centre of the metaphor, which relates back to the first sense. Note that they were more likely to give the second half of the proverb, *speech is silver*, as a continuation because in German both parts of the proverb are usually quoted together whereas in English *silence is golden* is much more frequent than *speech is silver*. Recall that English monolinguals created continuations for *silence is golden* that contrasted positive and negative aspects of keeping silent. While German monolinguals did focus mostly on the positive aspects of *silence*, note that, similar to the English monolinguals, they too contrast keeping silent as obeying a social norm versus keeping silent as a means to an end, e.g. to withhold information.

- (19a) *Wut ist ein Vulkan im Körper / von Emotionen.*
 Anger is a volcano in the body / of emotions.
 ‘Anger is a volcano inside a person / of emotions.’ (**emotionally**)
- (19b) *Wut ist ein Vulkan vor dem Ausbruch / kurz vor dem Ausbruch / mit Ausbruchsfahr / von dem man nie weiss, wann er ausbricht / mit zerstörerischer Macht / der entbrennt wenn er gereizt wird / auf dem man nicht spielen sollte.*
 Anger is a volcano before the eruption / shortly before the eruption / with eruption danger / of which one never knows when it will erupt / with destructive power/force / which inflames when it teased is / on which one not play should.
 ‘Anger is a volcano before eruption / just before/close to eruption / with a danger of eruption / where one never knows when it will erupt / with destructive power / which erupts when it is teased / which one should not play on.’ (**on the verge of eruption, holding back anger**)
- (19c) *Wut ist ein Vulkan im Ausbruch.*
 Anger is a volcano during eruption.
 ‘Anger is a volcano in action / an erupting volcano.’ (**erupting, venting anger**)

German monolinguals create continuations with the exact same three senses that English monolinguals implied in their continuations and often the phrasing of the senses is identical. Just as English monolinguals, the emotional sense is the one the continuations of German monolinguals allude to in the majority of cases. The other two senses are both part of the most frequent

general emotional senses, but they focus on the features that *anger* and *volcano* share, i.e. the salient features that constitute the metaphoric analogy between the metaphor source concept *volcano* and the target concept *anger*. When we compare continuations of all three senses to those of the English monolinguals, it is remarkable how similar the phrasings are, which would surely indicate very similar mental representations of the meaning of this cross-linguistically shared metaphor.

English-specific metaphors

- (20a) *Liebe ist eine Reise auf dem Meer des Lebens / in die Zukunft.*
 Love is a journey on the sea of the Life / in(to) the future.
 ‘Love is a journey on the sea of life / into the future.’ (**a part of life**)
- (20b) *Liebe ist eine Reise auf einer langen steinigen Straße / in das Ungewisse / mit Hindernissen / ohne Fahrplan / ohne Reiseführer.*
 Love is a journey on a long stony street/road / in(to) the uncertain / with obstacles / without an itinerary / without a tourist guide.
 ‘Love is a journey on a long rocky road / into the unknown / with obstacles / without an itinerary / without a tourist guide.’ (**ups and downs**)
- (20c) *Liebe ist eine Reise der Gefühle / ins Glück / zum vollendeten Glück.*
 Love is a journey of the emotions / into luck/happiness / (un)to complete luck/happiness.
 ‘Love is a journey of emotions / to happiness / unto complete happiness.’ (**emotionally**)
- (20d) *Liebe ist eine Reise in Gesellschaft eines anderen / mit jemand anderem durch's Leben / zu zweit.*
 Love is a journey in the company of someone else / with someone else through the life / as two together.
 ‘Love is a journey in the company of someone else / with someone else through life / together.’ (**personally**)

We know the intended meaning of *in* in *in die Zukunft* is *into the future* not *in the future* because for *in the future* the definite determiner would be *der*, *in der Zukunft* (*in the future*). The same is true for *in das Ungewisse*, which we know means *into the unknown* because if it was to mean *in the unknown*, it would have to be *im Ungewissen*. *Liebe mit Hindernissen* (*love with obstacles*) is a very common and popular idiom in German. Note, just like *tourist guide* in English can refer to either a person or a guide book, so does *Reiseführer* have the exact same two senses in German. Recall that English monolinguals alluded to wordings from wedding vows such as *through good times and bad* and talked about marriage; German speakers did not explicitly phrase their continuation this way, but the emotional, personal, and ups-and-downs senses could also appear in the specific context of wedding ceremonies, e.g. *zu zweit durch's Leben gehen* (*to walk through life together*, literally, *as two together*). In the literal gloss of *Glück* I have translated it as both luck and happiness. Recall the example I gave early on: while English has two words with distinct linguistic forms for the concepts ANIMAL and MEAT OF THAT ANIMAL, e.g. *cow* and *beef*, *pig* and *pork*, French does not have words with distinct linguistic forms, but uses the same for both concepts, e.g. *porc* for *pig* and *pork*. The

same applies to *Glück*. Because both the concept LUCK and HAPPINESS map onto the same linguistic form, the linguistic form, depending on its context is ambiguous or not between the two concepts. For instance, in the phrase *das Glück des Menschen* (*man's luck/happiness*), *Glück* is ambiguous between the two concepts and speakers could therefore also assume that it means both, i.e. happiness as serendipity, and this is the case for the continuation above. *Glück* could also be unambiguous in a context such as *da hast du aber nochmal Glück gehabt* (literally *you got lucky*, but idiomatically is akin to English *you got off easy*) where it clearly refers to the concept LUCK. Note that the first sense in the continuations, the part-of-life sense, can be taken to indicate that German monolinguals have made the conceptual connection between *love is a journey* and *life is a journey*, the same connection that was evident in the continuations created by English monolinguals. It would thus seem that with respect to this connection, as well as the other sense, English and German monolinguals seem to conceptually agree on the various aspects of the meaning of the metaphor.

German-specific metaphors

- (21a) *Sport ist Mord für unsportliche Leute / für Untrainierte / für Menschen ohne Übung / für Kinder, die gehänselt werden.*
 Sports is murder for non-athletic people / for the untrained / for people without practice / for children that teased are.
 ‘Sports is murder for the non-athletic / for the untrained / for people without practice / for children that are teased.’ (**the non-athletic, conventional sense**)
- (21b) *Sport ist Mord auf Raten.*
 Sports is murder on instalments.
 ‘Sports is murder in instalments (little by little).’ (**straining, exhausting**)
- (21c) *Sport ist Mord mit gesellschaftlicher Anerkennung / mit Prinzip / mit Vorzügen.*
 Sports is murder with societal acknowledgement / with principle / with advantages/as a prerogative.
 ‘Sports is murder with social prestige (acknowledgement in society) / with rules / with advantages.’ (**institutional**)

None of the German monolinguals created continuations that implied they took *murder* literally. Recall that this was a common interpretation made by English monolinguals. Another common interpretation by English monolinguals that took *murder* as a euphemism for *defeat* being the main objective in a competitive sport, may be a possible consequence for those German speakers who created continuations with the institutional sense, but it is certainly not their central focus. *Mit Prinzip* (*with principle*) refers to the rules and guidelines of a game or competitive sport because it occurs without a determiner, not the sense that *murder* is a euphemism for *defeat* as the main objective or the guiding principle, in which case the noun phrase would have to contain an indefinite determiner to invoke this sense, *mit einem Prinzip* (*with a principle*), or would have to be plural, *mit Prinzipien* (*with principles*, as in English *on principle*). The overwhelming majority of German speakers created continuations with the first sense, (21a). Note that this goes beyond the sense that physical exercise can be straining and exhausting, which was a prominent interpretation among English monolinguals, but makes a categorical distinction between those who have a natural talent and interest in sports and those who are not: *unsportliche Leute* (*non-athletic people*) are not just those who are not in shape to do

well in a particular sport—although that is certainly part of the idiomatic sense—but they are people who are non-athletic *because* they do not feel the need to be athletic; they are the ones who do not *want* to be athletic and what makes it difficult for them is the social pressure to do well. And recall that the only English speakers whose continuation in (13b) came closest to the German idiomatic meaning was a bilingual (English L1) speaker. The idiomatic meaning of the German-specific metaphor is the sense that German monolinguals allude to most in their continuations and *although* the other senses are conceptually plausible (those focussed on physical exercise being straining and exhausting and the nature of organised competitive sports) they are far less common than the idiomatic sense.

- (22a) *Der Teufel ist ein Eichhörnchen mit einem Talent für unerwartete Sprünge*
 The devil is a squirrel with a talent for unexpected jumps
 / *mit bösen Absichten.*
 / with evil intentions.
 ‘The devil is a squirrel with a talent for unexpected jumps / with evil intentions.’
 (sudden misfortune)

- (22b) *Der Teufel ist ein Eichhörnchen im Teufelsgewand / in einem*
 The devil is a squirrel in the devil’s clothes / in a
Wolfskostüm / mit falschem Gesicht / mit Hörnern / aus der
 wolf costume / with (a) false face / with horns / out of the
Hölle / mit glühenden Augen.
 hell / with glowing eyes.
 ‘The devil is a squirrel disguised / looking like the devil / in disguise / with horns / from hell / with glowing eyes.’ (the devil disguised as a squirrel)

Recall that the English monolinguals when they interpreted the metaphor in the sense of misfortune they focussed primarily on squirrels stealing nuts. When the German speakers create continuations alluding to this sense they specifically focus on that misfortune being sudden and unexpected or as an omen (*with evil intentions*), they do not describe the squirrel simply as a nuisance. The compound noun with *Gewand* (archaic *clothes, robes*) as the second noun is a common idiom that means *disguised as* whatever the first noun in the compound denotes. *Im Wolfskostüm* (*in a wolf costume*) is a common variation of another idiom that implies a disguise, *ein Wolf im Schafspelz* (*a wolf disguised as a sheep, wearing the sheep’s fur*). When the German speakers take the metaphor in the sense of the devil being disguised as a squirrel their continuations are conceptually the same that the English speakers gave: they either explicitly say *disguised* or use other idioms that imply a *disguise* or they attribute the devil’s outward appearance to the squirrel (*with horns, with glowing eyes*).

Novel metaphors

- (23a) *Ein Zuhause ist Geld auf das man bauen kann / auf der Bank / für*
 A home is money on that one build can / on the bank / for
die Zukunft / aus Stein / in Immobilienform / in der Rezession / in
 the future / out of stone / in real estate form / in the recession / in
stabiler Form / zur Seite gelegt.
 stable form / to the side put.
 ‘A home is money that one can rely on / in the bank / for the future / out of stone / on the property market / during recession / in stable form / put aside.’ (an investment in the future)

- (23b) *Ein Zuhause ist Geld zum Leben haben / für die Seele / mit
 A home is money in order for living have / for the soul / with
 einem schützendem Dach.
 a protective roof.
 ‘A home is money to live on / for the soul / with a protective roof.’ (**personally**)*

Very similar to English monolinguals, most of the continuations created by German monolinguals for this metaphor alluded to a home as an investment in one’s personal future. Recall that English monolinguals created continuations with the sense that a home is costly which none of the German speakers did. In *in stabiler Form (in stable form)*, *stable* could be used in one of two senses: (1) architecturally stable, and (2) financially stable. Note that it is *zum Leben haben (to live on)*, not *für’s Leben (for life)*: the focus is that a home is a foundation for a secure life and lifestyle, not that a home, assuming we are thinking of a house, is the kind of investment that lasts a lifetime. Hence the focus is on the emotional safety it provides and does not see it primarily as a financial investment in this case, but we can also see how it relates to the first sense.

- (24a) *Ein Freund ist ein Magnet für Freude und Zweifel / für all deine Gefühle
 A friend is a magnet for fun and doubts / for all your feelings
 / für das Glück / zum Glück / zu dem man sich
 / for the luck/happiness / to the luck/happiness / to whom one-self
 hingezogen fühlt / zur inneren Balance / um die Lücken in
 attracted-to feels / to inner balance / in order to the gaps in
 einem zu füllen.
 one(self) to fill.
 ‘A friend is a magnet for fun and doubts / for all your feelings / for happiness / to
 happiness / that you feel attracted to / in order to fill the void in oneself.’ (**emotionally**)*

- (24b) *Ein Freund ist ein Magnet in orientierungslosen Zeiten.
 A friend is a magnet in without-orientation times.
 ‘A friend is a magnet in times without orientation.’ (**morally**)*

- (24c) *Ein Freund ist ein Magnet für den anderen Freund / auf den man
 A friend is a magnet for the other friend / on whom one
 wieder zurückkommt / für mehr Freunde / zu noch mehreren.
 again returns-to / for more friends / to yet more.
 ‘A friend is a magnet for the other friend / to return to / for more friends / for more.’
 (**interpersonally**)*

We see the same three senses as the ones in the continuations of English speakers. Note that I have put *zu dem man sich hingezogen fühlt (that you feel attracted to)* with the emotional sense because the reflexive pronoun *sich* in conjunction with the verbs *hingezogen fühlt (feel attracted to)* means emotional attraction between people, not magnetic attraction in the physical sense because in that case a construction without the verb *fühlen (to feel)* and without the reflexive pronoun would be used, e.g. *der einen anzieht (that, meaning the magnet, is attracting one)*. When we look at the continuations of all three senses and compare them to those created by English monolinguals, we see that the wording is very similar as well as being conceptually similar.

The pattern that emerges from comparing the continuations created by monolinguals is thus as follows: the continuations are similar in terms of their linguistic form, the senses they allude

to, and the conceptually salient features of the concepts involved, whenever the metaphors are cross-linguistically shared or novel. For language-specific metaphors we see that native speakers of the two languages diverge for some of the senses but for others they agree. The senses that differ are those that allude to the conventional idiomatic meaning: native monolingual speakers create continuations that allude to the idiomatic meaning while monolinguals of the L2 do not. The senses that monolingual speakers of both languages come up with for the language-specific metaphors are senses that are conceptually plausible and built from concepts that in the previous chapter we saw would be available to speakers of both English and German. It thus seems that whenever conventional idiomatic meanings are unknown to speakers they rely on conceptually plausible inferences based on conceptually salient features of metaphor source and target concept as well as the implied metaphoric analogy. This finding agrees with the ratings given in the plausibility judgement task in part 1 of the experiments, where monolinguals rated L2 language-specific metaphors significantly lower than L1 and cross-linguistically shared metaphors. The continuations thus corroborate our conclusion in part 1 that the idiomatic meaning of L2 language-specific metaphors is not cross-linguistically intelligible and we can add that it is only conceptually salient meanings, where speakers of both languages have the same inventory of concepts and conceptually plausible mappings, that match cross-linguistically. If this result is because monolinguals are unaware of the idiomatic meanings of L2 metaphors, then bilinguals, who know these idiomatic meanings, should create continuations that allude to them.

Continuations created by bilinguals

Cross-linguistically shared metaphors

- (25a) *Unwissenheit ist ein Segen für die Menschheit / denn Wissen kann Ignorance is a blessing for the mankind / because knowledge can gefährlich sein. dangerous be.*
 ‘Ignorance is bliss for mankind / because knowledge can be dangerous.’ (**generic**)
- (25b) *Unwissenheit ist ein Segen für den sich Sorgenden / für die Ängstlichen Ignorance is a blessing for the one who worries / for the timid / für den Mutlosen. / for the dispirited.*
 ‘Ignorance is bliss for the one who worries / for the timid / for the dispirited.’ (**state of mind**)
- (25c) *Unwissenheit ist ein Segen für alle, die dumm sind / für den Dummen Ignorance is a blessing for all that dumb are / for the dumb / für den Unwissenden / für den Unbedarften / für den Ahnungslosen. / for the unknowing / for the ingenious one / for the clueless.*
 ‘Ignorance is bliss for all ignorant ones / for the ignorant / for the unknowing / for the ingenious one / for the clueless.’ (**naivety**)
- (25d) *Unwissenheit ist ein Segen für kurze Zeit / von kurzer Dauer / Ignorance is a blessing for a short time / of short duration / solange man ein Kind ist. as-long-as one a child is.*
 ‘Ignorance is bliss for a short time / of short duration / while you are a child.’ (**of limited duration**)

- (25e) *Unwissenheit ist ein Segen mit Haken / ohne Freude / von Ignorance is a blessing with a catch / without fun / of trügerischem Wert / mit Folgen.*
deceptive value / with consequences.
'Ignorance is bliss with a catch / without fun / of deceptive value / with consequences.'
(**false solace**)

These are the same five senses as the ones given by English and German monolinguals. Conventional senses are most frequent. Second most frequent are conceptually plausible senses. *There's a catch*, in (25e), is the idiomatic equivalent of the German expression *da ist ein Haken dran* (literally *there's a hook on there*). Notice the idiomatic expressions in both languages refer to the same reference frame: the hook on a fishing rod is the focus of the German idiom and the fish is caught with the hook, the fish being the catch that is the focus of the English idiom.

- (26a) *Schweigen ist Gold zur rechten Zeit / im richtigen Moment / in verwickelten Situationen / in bestimmten Situationen zumindest / zum Bewahren der Diskretion.*
Silence is gold at the right time / in the right moment / in tricky situations / in certain situations at least / in order to keep the discretion.
'Silence is golden at the right time / in the right moment / in tricky situations / at least in certain situations / in order to be discrete.'
(**institutional**)

- (26b) *Schweigen ist Gold und Reden ist Silber.*
Silence is gold and speech is silver.
'Silence is golden and speech is silver.'
(**conventional**)

- (26c) *Schweigen ist Gold zum Erhalten der Freundschaft.*
Silence is gold in order to sustain the friendship.
'Silence is golden in order to sustain friendships.'
(**interpersonally and personal needs**)

- (26d) *Schweigen ist Gold für Geheimniskrämer / für Leute, die etwas zu verbergen haben / für den Wissenden / für den, der zu viel weiß.*
Silence is gold for the mystery-monger / for people who something to hide have / for the knowing / for the one who too much knows.
'Silence is golden for the mystery-monger / for people who have something to hide / for one who knows / for the one who knows too much.'
(**withholding knowledge is power**)

- (26e) *Schweigen ist Gold ohne Wert.*
Silence is gold without value.
'Silence is gold without value.'
(**no actual gold**)

Bilinguals gave exactly the same senses as English and German monolinguals. *Geheimniskrämer* (*mystery-monger*), in (26d), is a common idiomatic compound noun in German. The translation is the established conventional-literal translation of the compound noun. Notice that the continuation *für Leute, die etwas zu verbergen haben* (*for people who something to hide have*) is simply a paraphrase of the concept *mystery-monger*. Similar to German monolinguals, bilinguals were a bit more likely than English monolinguals to give the second half of the proverb, *Reden ist Silber* (*speech is silver*), as continuation.

- (27a) *Wut ist ein Vulkan den es zu kontrollieren gilt / binnen kurzer Zeit
 Anger is a volcano that it to control is / within short time
 bricht sie aus und ist nicht mehr zu stoppen / von dem man
 breaks it (anger) out and is not anymore to stop / of that one
 nie weiß wann er/sie ausbricht / der Rache / auf wackligen Boden
 never knows when it erupts / of the revenge / on shaky ground
 / impulsiv und vernichtend.
 / impulsive and destructive.*

‘Anger is a volcano that needs to be controlled / within a short time it (anger) erupts and cannot be stopped / that one never knows when it will erupt / of revenge / on shaky ground / impulsive and destructive.’ (**emotionally**)

- (27b) *Wut ist ein Vulkan vor dem Ausbruch / der jederzeit ausbrechen
 Anger is a volcano before the eruption / that any time erupt
 kann / mit zerstörerischer Kraft / mit Sprengkraft.
 can / with destructive power / with explosive force.*

‘Anger is a volcano before eruption / which can erupt at any time / with destructive power / with explosive force.’ (**on the verge of eruption, holding back anger**)

- (27c) *Wut ist ein Vulkan während seines Ausbruchs / vor dem man
 Anger is a volcano during its eruption / before which one
 flüchtet / mit einer Explosion.
 flees / with an explosion.*

‘Anger is a volcano during its eruption / that you run away from / with an explosion.’ (**erupting, venting anger**)

den es zu kontrollieren gilt must mean emotionally because there is no way to control an actual volcano; maybe one can monitor its activity and even predict an eruption, but there is no way to prevent or control it. *von dem man nie weiß wann er/sie ausbricht* (that one never knows when it will erupt) anaphoric pronoun varies between *er* referring to the volcano and *sie* referring to anger.

English-specific metaphors

- (28a) *Liebe ist eine Reise auf dem Lebensweg / durch mein Leben.
 Love is a journey on the life path / through my life.
 ‘Love is a journey on the path of life / through my life.’ (a part of life)*

- (28b) *Liebe ist eine Reise in die Ungewissheit / ins Ungewisse /
 Love is a journey in(to) the uncertainty / into the uncertain /
 ins Unbekannte / in das Verderben / auf unbestimmte Zeit / mit
 into the unknown / in(to) the perdition / for indefinite time / with
 Hindernissen / mit ungewissem Ausgang / mit unbekanntem Ziel.
 obstacles / with uncertain outcome / with unknown end.
 ‘Love is a journey into uncertainty / into the unknown / to perdition / for an indefinite period of time / with obstacles / with uncertain outcome / to an unknown end.’ (**ups and downs**)*

- (28c) *Liebe ist eine Reise ins Glück / zum Glück /
 Love is a journey into the luck/happiness / to the luck/happiness /
 zum Gipfel des Glückes / auf dem Meer des Glücks
 onto the peak of the luck/happiness / on the sea of the luck/happiness
 / in das Land der Gefühle.
 / in(to) the land of the emotions.*

‘Love is a journey to happiness / towards happiness / onto the peak of happiness / on the sea of happiness / into the land of emotion.’ (**emotionally**)

- (28d) *Liebe ist eine Reise zu zweit / mit dem Partner / von Herz*
 Love is a journey as two together / with the partner / from heart
zu Herz / zum Herzen.
 to heart / to the heart.

‘Love is a journey together / with a partner / from one heart to the other / to the heart.’
 (**interpersonally**)

Similar to German monolinguals, English-German bilinguals mention all the senses that English monolinguals mention except the reference to wedding vows: again, many of the phrases they do use in their continuation would also come up at weddings, *zu zweit*, *auf dem Lebensweg*, *zum Herzen* (*to the heart*) meaning the heart of the other person, so interpersonal. Notice the similar mental image bilinguals have here, *der Gipfel des Glückes* (*the peak of happiness*), similar to English *her happiness was of paramount importance to him*. Note it says *eine Reise zum Herzen* (*a journey to the heart*) which must refer to the heart of the person that is the object of one’s desire, not *eine Reise des Herzens* (*a journey of the heart*), in which case it would refer to one’s own heart and thus be a statement about the emotions associated with love. It thus seems that bilinguals’ mental representation of this metaphor is closer to English monolinguals and that we only find those senses match between all monolinguals and bilinguals that are conceptually salient regardless of the idiomatic meaning.

German-specific metaphors

- (29a) *Sport ist Mord für den Unsportlichen / aber nur für die Unsportlichen*
 Sports is murder for the non-athletic / but only for the non-athletic
/ die ihn nicht mögen / von Untrainierten.
 / those that it (sports) not like / of the untrained.

‘Sports is murder for the non-athletic / but only for the non-athletic / those that don’t like it / of the untrained.’ (**the non-athletic, conventional sense**)

- (29b) *Sport ist Mord am Körper / an der eigenen Gesundheit / trotz*
 Sports is murder on the body / on the own health / despite
Vorsicht / auf Raten / bei falscher Dosierung / in ungesundem
 caution / on instalments / in the wrong dosage / in unhealthy
Maß betrieben / bei Überanstrengung / bei Übertreibung /
 amount practised / in case of overexertion / in case of over-doing it /
wenn man es übertreibt.
 when one it over-does.

‘Sports is murder for the body / of one’s health / despite caution / in instalments (little by little) / in the wrong dosage / practised in unhealthy amounts / in case of overexertion / in case you over-do it / when you over do it.’ (**straining, exhausting**)

- (29c) *Sport ist Mord im Leistungssport / mit Freunden.*
 Sports is murder in competitive sports / with friends.

‘Sports is murder in competitive sports / with friends.’ (**institutional**)

Bilinguals, similar to German monolinguals, create continuations that allude to the conventional idiomatic meaning. The conventional sense is alluded to most often in the continuations. The second-most common sense is the conceptually plausible straining/exhausting sense that is most common in the continuations created by English monolinguals.

(30a) *Der Teufel ist ein Eichhörnchen der unerwartet Sprünge / mit bösen Gedanken / mit schlechten Intentionen / genauso hinterlistig.*
 The devil is a squirrel of the unexpected jumps / with evil thoughts / with bad intentions / equally insidious.
 ‘The devil is a squirrel of unexpected jumps / with evil thoughts / with bad intentions / equally insidious.’ (**sudden misfortune**)

(30b) *Der Teufel ist ein Eichhörnchen mit Hörnern / mit Pferdefuss / im Kostüm / im Schafspelz / mit drei goldenen Haaren / mit rotem Fell.*
 The devil is a squirrel with horns / with a horse foot / in the costume / in the sheep skin / with three golden hairs / with red fur.
 ‘The devil is a squirrel with horns / the foot of a horse / in disguise / with three golden hairs / with red fur.’ (**the devil disguised as a squirrel**)

Im Schafspelz (literally *in a sheep skin*) is a common idiom for *disguise*. Notice that similar to the German monolinguals, bilinguals did not create any continuations that implied that the squirrel is merely a nuisance. For German monolinguals I suggested that we take this as an indicator that they are aware that the sense ‘squirrel as a nuisance’ would conflict with the idiomatic meaning and for this they would have to be aware of the idiomatic meaning in the first place. Hence, I suggest that we conclude that the bilinguals are also aware of the same idiomatic meaning that the German monolinguals are aware of.

Novel metaphors

(31a) *Ein Zuhause ist Geld für die Zukunft / fürs Alter / gut angelegt*
 A home is money for the future / for the old age / well invested
 / *das gut angelegt ist / bei dem man sicher sein kann, es ist gut*
 / that well invested is / where one sure be can it is well
angelegt / auf einem sicheren Platz.
 invested / on a safe place.
 ‘A home is money for the future / for old age / well invested / that is well invested / where you can be sure it is well invested / in a safe place.’ (**an investment in the future**)

(31b) *Ein Zuhause ist Geld doch es kostet auch / ohne Nutzen / wenn man es hat / wenn man es verkauft.*
 A home is money, but it costs also / without utility / if one it has / if one it sells.
 ‘A home is money, but it’s also expensive / that you can’t spend / if you have it (the money) / if you can sell it (the home).’ (**costly**)

(31c) *Ein Zuhause ist Geld und noch viel mehr wert / eine Familie ein Schatz / mit Freude / mit Wohlbehagen / mit Zufriedenheit gepaart.*
 A home is money and yet much more worth / a family a treasure / with happiness / with feeling comfortable / with satisfaction paired.
 ‘A home is money and yet worth so much more / (and) a family (is) a treasure / with happiness / that makes you feel comfortable / paired with satisfaction.’ (**personally**)

Note that bilinguals similar to English monolinguals and unlike German monolinguals created continuations that implied that a home could also be costly. Notice that while the first two

senses focus more on the monetary value, the fact that bilinguals created continuations that allude to the last sense indicates that bilinguals, similar to monolinguals, are able to infer a metaphoric meaning. So while plausibility ratings are on the low end of the ratings scale, speakers are still able to create a context that will boost plausibility, which presupposes that there is something about the novel metaphor *a home is money*, in the metaphor source and target concept, that allows speakers to extend upon it conceptually.

(32a) *Ein Freund ist ein Magnet für viele Freuden / in einsamen Zeiten.*
 A friend is a magnet for many joys / in lonely times.
 ‘A friend is a magnet for many joys / in lonely times.’ (**emotionally**)

(32b) *Ein Freund ist ein Magnet um Halt zu bekommen /
 A friend is a magnet in order to security receive / that
 der immer hilfsbereit ist.*
 always willing to help is.
 ‘A friend is a magnet that gives you security / helps you stay grounded / that is always willing to help.’ (**morally**)

(32c) *Ein Freund ist ein Magnet für den anderen / zu dem es dich
 A friend is a magnet for the other one / to which one
 hinzieht / der viele weitere Freunde anzieht / für neue Freunde /
 is attracted / which many more friends attracts / for new friends /
 für seine Freunde / auch Gegensätze ziehen sich an / der dich selten
 for his friends / also opposites attract / which you seldom
 abstößt / der für einen da ist / der nie verloren geht / zum
 repels / who for one there is / who never lost goes / in order to
 Bleiben / zum Reden.
 stay / in order to talk.
 ‘A friend is a magnet for the other one / to which one is attracted / who attracts many more friends / for new friends / for his friends / (because) opposites attract / who rarely blows you off / who is there for you / who never goes missing/disappears on you / is there to stay / is (always) there to talk.’ (**interpersonally**)*

Halt geben has the same moral connotations as the English idiom *to stay on a straight path*. Recall that some English monolinguals create continuations along the lines of *to keep you on the right path*, where *right*, i.e. morally right, can be understood in the sense of *straight*, making *the straight path* semantically equivalent to *the right path*. *Help* in German *hilfsbereit* is *help* in the sense of a moral obligation. *Hilfsbereit* literally consists of *Hilfe* (*help*) and *bereit* (*willing*) and thus means *willing to help*. The corresponding noun, *Hilfsbereitschaft* (*the willingness to help*), is also conceptually the label for a virtue, thus showing again the moral impetus invoked by the continuations. This shows that while the metaphor received plausibility ratings on the low end of the rating scale, bilinguals, similar to monolinguals, are able to construct a context that extends the conceptual potential that must already be given in the metaphor source and target concept. Otherwise they would take the metaphor at face value, which is not what they are doing when they extend the moral connotations of friendship because the source concept *magnet*, at face value, has no moral connotations.

When we compare the continuations created by English and German monolinguals we see the following: they create similar continuations for cross-linguistically shared metaphors that imply the same sets of senses and they are primarily conventional idiomatic senses and secondarily

conceptually plausible senses. When they create continuations for language-specific metaphors, only the continuations of native speakers allude to the conventional idiomatic sense of the metaphor while non-native speakers create continuations that are conceptually sensible but which diverge from the idiomatic meaning. For novel metaphors speakers of both languages create continuations that are conceptually plausible and we find that they come to the same conceptually plausible senses. We can thus conclude two things: (1) whenever conventional idiomatic meanings are known to speakers (as in the case of cross-linguistically shared and L1 language-specific metaphors), speakers prefer them over meanings that are just conceptually plausible, and (2) when speakers do not associate any conventional idiomatic meaning with a particular linguistic metaphor (as in the case of novel and L2 language-specific metaphors), they rely primarily on conceptually plausible interpretations. We can therefore conclude that whenever conventional linguistic meanings are available to speakers, they take precedence over, i.e. they are more salient than pure considerations of conceptual plausibility. We can thus add to our answer of question 1 that L2 language-specific metaphors are cross-linguistically intelligible between the two languages in close contact only in the sense that non-native speakers come to the same conceptual interpretations as native speakers, but non-native speakers are *unable* to infer those meanings that native speakers consider conventional idiomatic meanings.

The same line of reasoning holds true for bilinguals: (1) they know conventional idiomatic meanings for cross-linguistically shared as well as all language-specific metaphors and so these are the meanings that their continuations allude to most often, more often than continuations that allude to merely conceptually plausible inferences. (2) For novel metaphors, bilinguals are in the same position as monolinguals: they do not know of any conventional meanings, hence their inferences focus on conceptually salient features of metaphor source and target concepts and on drawing further analogies from the implied metaphoric analogy. Therefore, bilinguals come up with continuations for novel metaphors that are very similar to those created by monolinguals and we see that the range of senses they allude to is limited to the same set of concepts and conceptual mappings that monolinguals used to construct continuations for the novel metaphors. This suggests that, apart from bilinguals' knowledge of idiomatic meanings for the L2 metaphors, bilinguals seem to make use of the same inventory of concepts and conceptual mappings as the English and German monolinguals when they make inferences about the metaphors. It would thus seem that monolinguals' lower plausibility ratings for L2 metaphors could be solely attributed to them being unfamiliar with the necessary idiomatic meanings. Bilinguals, on the other hand, who also give lower ratings to L2 metaphors seem to do so as an expression of their awareness of metaphors' language-specificity because their continuations show an awareness of the idiomatic meanings and that awareness was also evident in post-experimental interviews.

We might say that for metaphoric proverbs where speakers know idiomatic meanings, they have very high expectations as to what constitutes a plausible context. On this relevance-theoretic explanation, it makes sense that monolingual native speakers should perceive L1-specific and cross-linguistically shared proverbs to be less plausible in context than out of context. With respect to bilinguals, it makes sense why the effect of context on plausibility judgements should be greater for L1-specific proverbs than shared ones because bilinguals are aware of and sensitive to the language specificity of the proverbs; part of their mental representation would be what target languages apply for a metaphoric proverb. Monolinguals, on the other hand, do not have

this sensitivity; they are not even aware, as post-experimental interviews revealed, that the L1 metaphors are only proverbs in their native but not the other language. Why the effect of context on monolinguals' plausibility judgements is greater for L1 than shared proverbs remains an open question, but whatever the reason is, it means that speakers' plausibility judgements, monolinguals' and bilinguals' alike, are not simply an expression of conceptually motivated intuitions. We can therefore not assume that plausibility judgements provide us with a direct insight into the mental representation of conceptual knowledge; rather the impression we get is mitigated and filtered through linguistic constraints. For English monolinguals, the effect of context on metaphor plausibility is almost twice as strong as it is for German monolinguals, and for German monolinguals, plausibility of L1 proverbs decreases by as much as it decreases for shared proverbs. I therefore think it would be interesting to explore in future research whether there is something about the particular English proverbs that causes this or whether it is a sampling effect of the number of participants (recall, finding monolingual German native speakers who do not speak English is exceedingly and increasingly difficult in present-day Germany due to the fact that English is, de facto, a compulsory second language in the secondary and tertiary education system).

7.5 Summary

CMT claims that idiomatic meanings of metaphors are salient because of the way that concepts are intertwined into ever more complex higher-order structures in speakers' conceptual systems through conceptual metaphors which yield the corresponding inferences. Conceptual metaphors are associations between concepts that form when these associations are drawn frequently. As such, CMT claims that conceptual metaphors are what underlie meaning salience. According to CMT, the systematicity of metaphors in linguistic expression gives insight into the kinds of conceptual metaphors. RT and GS, contrary to CMT, claim that while meaning salience may in part depend on the mental representation of concepts, the systematicity of metaphors in linguistic expression cannot be taken as *direct* evidence of conceptual metaphors and metaphor comprehension involves *non-conceptual* linguistic conventions which may take precedence over purely conceptual intuitions. In this chapter, we looked at the question: Are metaphors from one language intelligible to speakers of another language if those two languages are in close linguistic contact and thus can be assumed to share similar or the same concepts?

We defined cross-linguistic conceptual similarity in terms of metaphors' lexical meanings, their commonness in use, speakers' familiarity with the metaphors, similar expectations of contextual relevance, and related metaphoric imagery. If CMT is correct in its claims that metaphor comprehension is a matter of conception, then when we pick language-specific metaphors such that the source and target concepts of these metaphors are cross-linguistically approximately similar between two languages, these metaphors should be cross-linguistically communicative to speakers of either language. If, however, RT and GS are correct, then we should find that cross-linguistic communicability of L2 language-specific metaphors is lowered for native speakers who do not speak the other language (*monolingual*), even when cross-linguistic approximate conceptual similarity is maximised.

If L2 language-specific metaphors are cross-linguistically communicative, then they should be read as fast as L1-specific and cross-linguistically shared metaphors (*strong communicability*) or at least faster than unfamiliar novel metaphors (*weak communicability*). Processing delays

correlate with processing difficulty, i.e. the more complex inferential processing of a particular metaphor is, the longer it will take. For L1-specific and cross-linguistically shared metaphors, monolingual speakers know their idiomatic meanings. These meanings serve as inferential shortcuts. There are no established idiomatic meanings for the novel metaphors. Since there are no inferential shortcuts available for novel metaphors, processing should be delayed. Experimental results show that monolinguals' reading/response times are significantly slower for L2 language-specific metaphors than L1 and shared metaphors and as slow as for novel metaphors. Measured through reading/response times, L2-specific metaphors are not cross-linguistically communicative, neither in the weak nor the strong sense.

In order for a metaphor to be communicative, it must first be intelligible, and in order to be intelligible, the analogy implied between its source and target concept must be plausible. Metaphor plausibility should thus be a prerequisite for communicability. If L2 language-specific metaphors are cross-linguistically communicative, then monolingual native speakers should find them as plausible as L1-specific and cross-linguistically shared metaphors (*strong communicability*) or at least more plausible than unfamiliar novel metaphors (*weak communicability*). Experimental results show that English and German monolinguals find L2 language-specific metaphors significantly less plausible than L1-specific and cross-linguistically shared metaphors, but more plausible than novel metaphors. Measured through plausibility judgements, L2 language-specific metaphors are only cross-linguistically communicative in the weak but not the strong sense.

If the meaning that native speakers conventionally infer as a salient meaning for a metaphor can be thought of as the context omitted from overt linguistic expression that would be necessary to 'spell out' that meaning, then L2 language-specific metaphors are cross-linguistically communicative if non-native speakers are able to 'spell out' contexts similar to the ones that native speakers 'spell out.' Experimental results show that monolinguals create contexts for L2-specific metaphors that are different from those that native speakers create in order to allude to the conventional idiomatic meanings. However, the contexts that non-native speakers create for L2-specific metaphors are cross-linguistically similar to contexts that native speakers create whenever they do not allude to the idiomatic meanings, which shows that speakers of both languages do conceptualise similarly. The context creation task thus gives evidence that cross-linguistic approximate conceptual similarity is indeed maximised and that the lowered cross-linguistic communicability of L2 metaphors found in reading/response times and plausibility judgements is unlikely to be the result of differences in conception. Together, the experimental evidence suggests that speakers do not infer those meanings for L2 language-specific metaphors that native speakers consider to be the most salient meanings, but that it is not a significant difference in conception that prevents them from doing so—a finding which contradicts CMT, but supports RT and GS. Context is well-known to have a significant effect on inferential meaning. In the next chapter we will therefore investigate whether contexts that native speakers consider sensible and helpful can improve L2 metaphor communicability.

Chapter 8

Experiments: cross-linguistic metaphor intelligibility in context

8.1 Introduction

In Section (4.2) in Chapter 4, I introduced the two central questions of this research, the second of which will be the subject of this chapter: If context improves the intelligibility of metaphors across languages, do bilinguals make use of context in the same way that monolinguals do? In the previous chapter, we saw that when monolinguals see the metaphors out of context, they showed lower reading/response/reaction times and lower plausibility ratings for L2 than L1 or cross-linguistically shared metaphors. We should thus be interested in whether context could make the L2 metaphors more intelligible to monolinguals. In the previous chapter, we also saw that English-German bilinguals, who in post-experimental interviews demonstrated they were familiar with both L1 and L2 language-specific metaphors, read L2 English-specific metaphors as fast as L1 metaphors and not significantly slower than cross-linguistically shared metaphors. Bilinguals also found L2 metaphors as plausible as L1 and cross-linguistically shared metaphors. In this chapter we will see that when speakers see the metaphors in context, monolinguals find the L2 metaphors more intelligible than when presented out of context. Context has not significantly affected bilinguals' rating behaviour. Context must therefore provide monolinguals, but not bilinguals, with information which facilitates their performance during L2 metaphor comprehension. A potential explanation would be that bilinguals may ignore L2 metaphors' context because the information it provides is *not new* to them, whereas it is *new* to monolinguals.

Another important reason why we should be interested in the effect that context has on cross-linguistic metaphor plausibility is that out-of-context presentation of metaphors is not as *naturalistic* as in-context presentation. If we do not embed metaphors in context, our experiments are less realistic than the way that speakers would normally encounter them and consequently the experimental results would be less representative. The contexts we give to participants should thus be as natural as possible and that means that they should be designed such that facilitates them effective communication by increasing cross-linguistic intelligibility whenever possible. Recall that to this end, participants in the out-of-context experiments were asked to create contexts for eight of the metaphors that would make them more sensible. What

participants did not know is that these contexts would be part of the material given to the next groups of English monolinguals, German monolinguals, and English-German bilinguals. Although in-context metaphor presentation is more naturalistic, it is not the case that there was nothing insightful about metaphors' plausibility ratings out of context. Rather it is the way that monolinguals and bilinguals perceive metaphors' plausibility differently when they see them out of context than when they see them in context that reveals some important clues about how participants approach the judgement task and how cross-linguistically intelligible idiomatic meanings are.

Novel and shared metaphor plausibility ratings are not *ideal* control conditions because the ratings do not remain constant but change under the influence of context. Recall that in Section (4.4.1) in Chapter 4, I said that we need to define cross-linguistic intelligibility as a *relative measure*, one that looks at how judgements change depending on the information available (experimental conditions, independent variables). In this chapter we will see that the reason why we can only define it as a relative measure is because the plausibility ratings of novel and shared metaphors, relative to which we define weak and strong cross-linguistic intelligibility, vary under the influence of context.

8.2 Predictions

Monolinguals

CMT predicts: Given the lower ratings of conceptual plausibility for L2 and novel metaphors in out-of-context metaphor presentation, ratings should go up with context.

The linguistic theories RT and GS predict: Ratings for L2 and novel metaphors should go up with context.

CMT, RT, and GS predict that context should aid comprehension of conventional metaphors. Ratings for L1 and shared metaphors should go down if the context is not what speakers expect, i.e. if it conflicts with speakers' relevance expectations, and otherwise remain the same.

Bilinguals

CMT predicts: Ratings for L1, L2, and novel metaphors should go up with context. Ratings for shared metaphors should remain the same with context.

The linguistic theories RT and GS predict: ratings for novel metaphors should go up with context. Ratings for L1, L2, and shared metaphors should go down if the context is not what speakers expect and otherwise remain the same and there should be no change if contextual information is already included in the arbitrary linguistic meaning (sense relations) that speakers have.

8.3 Definitions and operationalisations

Recall the definitions of weak and strong cross-intelligibility, the reason for and wording of the plausibility judgement task, and the closeness in language contact between English and German developed in Chapter 4. In order for the results of in-context experiments to be comparable to those of the out-of-context experiments, we must adhere to the same definitions

and operationalisations of intelligibility, the judgement task, and the closeness of language contact. They are repeated here as a reminder. What we need to amend is an operationalisation of the contributions made by context.

- Intelligibility in the strong sense (same as in out-of-context experiments): Language-specific metaphors are perceived to be as conceptually plausible as metaphors that are cross-linguistically shared.
- Intelligibility in the weak sense (same as in out-of-context experiments): Language-specific metaphors are perceived to be conceptually more plausible than novel metaphors.

Task (same as in part 1 of out-of-context experiments): “How much sense does a metaphor make?” = conceptual plausibility as a measure of cross-linguistic intelligibility

Closeness of language contact between English and German (same as in out-of-context experiments)

Speaker groups in in-context experiments are the same three as in out-of-context experiments (English monolinguals, German monolinguals, and English-German bilinguals), but new speakers were recruited for all groups to ensure *independence of samples*.

CMT, RT, and GS agree that contextual information is relevant if it yields positive cognitive effects, is informative, contributes something new to the discourse, aids comprehension, and is an explanation of conceptual metaphoric mappings (i.e. it is a substitute for the conceptual metaphors that speakers might lack). More context should make it easier to form *ad hoc* concepts, i.e. context is relevant if it is helpful during comprehension. If contextual information is relevant for comprehending a metaphor, then (all) speakers should (always) make use of it. If contextual information is irrelevant for comprehending a metaphor or conceptually conflicts with what speakers already inferred, speakers should ignore this new contextual information. With relevant contextual information, cross-linguistically unintelligible metaphors should become intelligible.

All contexts in in-context experiments were designed to be relevant and helpful: for those created by native speakers in out-of-context experiments (monolinguals), speakers were instructed to make context helpful in understanding the metaphor, i.e. contribute to the metaphor’s meaning conceptually. Those created by the researcher were also created to be helpful. Some contexts will turn out to be suboptimal, but—importantly—not because they are conceptually “bad” but because they are not what speakers expected (linguistic convention, relevance expectation, salient context) and that expectation is language-specific.

8.4 Material and methods

8.4.1 Participants

In the experiments of the previous chapter, participants saw the metaphors out of context. In the experiments of this chapter, they see them in context. Fifty-nine English monolinguals (44 females; mean age = 26, SD = 5.7), 20 German monolinguals (10 females; mean age = 36, SD = 15.3), and 21 English-German bilinguals (10 females; mean age = 35, SD = 14.1) participated in the in-context experiments. Language proficiency (whether they qualify as monolingual native

speakers or as bilinguals) was recorded as the number of years of institutional instruction in the non-native language, the duration of time spent abroad, language proficiency of parents and frequency of use. Bilinguals include advanced second language learners with an extended stay or permanent residence in the L2 country. Classified as *German monolinguals* were speakers who had spent less than 6 months in an English-speaking country (most stated typical two-week vacations which added up, averaged out to 6 months) and less than a year of English as a second language; half of them did not have English as a second language *at all* because they were from former East Germany where their first second language was Russian (as indicated by them in their personal form). The average participant is 32 years old or younger. Note that English monolinguals are generally younger and vary less in age than the German monolinguals and the bilinguals because they are all university students whereas the German monolinguals and bilinguals include professionals as well as students. Most of the German monolinguals are from East Germany where they learned Russian as a second language at school, but never English.

8.4.2 Metaphor material

The experiments of this chapter use the same metaphor material as the out-of-context experiments of the previous chapter. However, new metaphors were added for better counterbalancing. Recall, in the out-of-context experiments, there were only 5 English-specific metaphors, but 12 German-specific ones. In the in-context experiments 7 English metaphors were added so that English and German metaphors are equally represented in the material with 12 metaphors each. One cross-linguistically shared and one novel metaphor were added. This brings the number of novel metaphors also to 12, the same as for English and German metaphors. With the additional cross-linguistically shared metaphor, this brings the total number of metaphors to 50 in the in-context condition. The additional metaphors were subjected to the same selection process as the metaphors in the out-of-context experiments. See Table (A.1) in the Appendix.

8.4.3 Context material

Recall that in part 2 of the out-of-context experiments in the previous chapter, participants were asked to create continuations to 8 of the metaphors. When I analysed their answers I sorted their continuations according to the sense that they alluded to. There were at least two senses alluded to by the continuation per metaphor and up to five senses. Hence, we need to decide which of the two to five senses we want to pick as the context that participants in the in-context experiments will see. I suggest that we select the one that occurs most often, which is either one with a conceptually plausible sense or the conventional idiomatic sense that speakers expect, including their conventional expectations of relevant contexts. More precisely, for cross-linguistically shared and language-specific metaphors, monolingual native speakers created continuations that most often alluded to the conventional sense. We therefore know to select the most general phrasing of a continuation that alludes to this sense. For novel metaphors, continuations most commonly alluded to a sense that was conceptually plausible and the less frequent other senses were, the less conceptually plausible they were. Hence, we should pick the most frequent continuations for novel metaphors because they are the conceptually most plausible. In the previous chapter, I already demonstrated that the conventional continuations for the cross-linguistically shared metaphors are cross-linguistically comparable between English and German. I also showed that the same is true for the conceptually most plausible continua-

tions for novel metaphors. Of course, we cannot make it so that the conventional continuations of language-specific metaphors are cross-linguistically comparable because that is part of our experimental investigation. We can, however, ensure that at the very least continuations for language-specific metaphors be conceptually plausible in a cross-linguistic sense. Recall that in the previous chapter I already demonstrated that monolinguals most frequently created conceptually plausible continuations for the L2 metaphors and that speakers of the L2, for which the metaphors were L1 metaphors, while they did create conventional continuations for the same metaphors, their second most common continuations were continuations that alluded to the same conceptually plausible sense as those created by monolinguals of the other language. We therefore want to pick these conceptually plausible continuations that monolinguals created as the context for L2 metaphors in the in-context experiments. For L1 metaphors, we could go with the continuation that is most likely to be conceptually plausible to L2 monolinguals and that comes closest to the conventional idiomatic meaning of the metaphor. Once we have singled out the appropriate continuations per metaphor type, we select the most common phrasing created by English and German monolinguals in order to ensure naturalness and cross-linguistic comparability of their translation. It makes sense to use those contexts that native speakers consider helpful in follow-up experiments which investigate whether context can help non-native speakers comprehend L2 metaphors similarly to how native speakers comprehend them.

Contexts picked from those created by the monolingual native speakers in the out-of-context experiments

- (33) *Unwissenheit ist ein Segen (für den Unwissenden).*
 Ignorance is a blessing (for the unknowing).
 ‘Ignorance is bliss (for the unknowing).’

This continuation that alludes to the sense labelled **naivety** in the previous chapter. Note that in the previous chapter, I have also listed other phrasings under this sense, such as *für die Dummen (for the dumb ones)*. I decided against *für die Dummen (for the dumb ones)* for two reasons: (1) it is unnecessarily negative, and (2) *Unwissenheit (ignorance)*, in both German and English, can be wilful or not, but *für die Dummen (for the dumb ones)* would only allude to one possible interpretation of *ignorance* while *for the unknowing* can apply to both. *For the unknowing* is therefore the more neutral and the more general continuation, which should therefore be conceptually most plausible cross-linguistically.

- (34) *Schweigen ist Gold (in Momenten der Andacht).*
 Silence is gold (in moments of solemnity).
 ‘Silence is golden (in a solemn moment).’

I said that for cross-linguistically shared metaphors, we want to pick those continuations that allude to the conventional sense and, indeed, the continuation most commonly created by German monolinguals *Reden ist Silber (speech is silver)* is the conventional second half of the proverb, the conventional sense, but it is not the most common continuation by English monolinguals. I therefore decided to go with the most common continuation created by English native speakers, which is the second most common one for German monolinguals, because this sense would still be very much conceptually plausible to speakers of both languages. The reason why *speech is silver* is only the second most common continuation by English speakers is easily

explainable when we look at the corpus frequency: The metaphor *silence is golden*, which is actually the second half of the proverb *speech is silver (and/but) silence is golden*, occurs much more often on its own than the proverb occurs as a whole. In the German corpora, the whole proverb *Reden ist Silber (und/aber) Schweigen ist Gold (speech is silver (and/but) silence is golden)* is much more frequent than in English. German speakers are thus more likely to be familiar with both parts of the proverb than English speakers. It therefore makes no sense to select *speech is silver* as a context in study #2 when English speakers are less likely to be familiar with it.

- (35) *Wut ist ein Vulkan (von Emotionen).*
 Anger is a volcano (of emotions).
 ‘Anger is a volcano (of emotion).’

The emotional sense was the most common sense that continuations by both English and German monolinguals alluded to and it is also the one that comes closest to the sense of the implied metaphoric analogy because alternative continuations, such as *before eruption* or *during eruption*, only allude to features of *volcano*, the metaphor source/vehicle concept. Note that the continuation *of emotion* includes the conceptual sense that a volcano of emotions can erupt, but this mental image of an eruption of emotion would be a metaphoric eruption. This eruption of emotions would not necessarily have to be in the form of anger; it could be a bursting into tears, out of sadness or overwhelming joy, or it could be agitation or nervousness. So while an eruption of emotions would not necessarily have to be negative, a volcanic eruption most certainly is negative, hence, this continuation comes closest to the sense of the metaphoric analogy.

- (36) *Liebe ist eine Reise (mit Hindernissen).*
 Love is a journey (with obstacles).
 ‘Love is a journey (with obstacles).’

For this English-specific metaphor, we choose the continuation that is (1) most common among English monolinguals and (2) conceptually most plausible to German monolinguals. In the previous chapter I labelled the sense that this continuation alludes to **ups and downs** because it recognises that a romantic relationship is not solely filled with joy. Other continuations that I put with this sense in the previous chapter include *auf einer langen steinigen Straße (on a long rocky road)* and *ins Ungewisse (into the unknown)* and *through good times and bad*, but *mit Hindernissen (with obstacles)* seemed to me the most general encapsulation of this. Also notice how *on a long rocky road* invokes the mental image of a specific type of journey, a road trip, but by using *with obstacles* as context we allow for other types of journeys as well while also incorporating the sense that on this journey there are ups and downs, good times and bad.

- (37) *Sport ist Mord (für unsportliche Leute).*
 Sports is murder (for non-athletic people).
 ‘Sports is murder (for the non-athletic).’

In the previous chapter I already discussed that while this continuation, which alludes to the conventional sense of the metaphor in German, is not the most common among English monolinguals, English speakers did allude to it in some of their continuations, which shows that it is a conceptually plausible interpretation to them. Keep in mind that we do not choose a phrasing

such as *the untrained* (*für Untrainierte*) or *for people without practice* (*für Leute ohne Übung*) because the idiomatic sense of the metaphor is that these are people who have no interest in sports, not just that they are bad at it, and these alternative phrasings might be misleading and too narrow. *Für unsportliche Leute* (*for the non-athletic*) can be read as both those who have no interest in sports, and as a result might also not be very good at it, or those who happen to be bad at it while they may or may not actually have an interest in being good at it.

- (38) *Der Teufel ist ein Eichhörnchen (mit bösen Absichten).*
 The devil is a squirrel (with bad/evil intentions).
 ‘The devil is a squirrel (with malice).’

I said in the previous chapter that none of the English monolinguals came up with a continuation that could be said to allude to the conventional idiomatic meaning of the German metaphor. The sense that comes closest to it conceptually, which also was the most frequent continuation among English monolinguals, was that of **squirrels as nuisance**. I therefore picked this continuation from among those created by the German speakers because due to its generality it can apply to both the devil and the squirrel and depending on which of the two it is applied to, what these bad intentions are is assigned a more specific meaning: in the case where it is applied to the squirrel, it yields the **squirrels as nuisance** sense that English monolinguals find conceptually most plausible. When it is applied to the devil, this gives rise to the appropriate extensions of *bad intentions* as *malice* and *evil intentions*. Since *malice* can therefore be applied to both the metaphor source and target concepts individually, English speakers should also be able to conceptually construe a representation of the implied metaphoric analogy where *malice* applies to both the devil and the squirrel.

- (39) *Ein Zuhause ist Geld (für die Zukunft).*
 A home is money (for the future).
 ‘A home is money (for the future).’

Seeing a home as **an investment in the future** was the most common sense that continuations created by English and German monolinguals alluded to. Other phrasings of continuations with this sense included the words *investment* or *real estate*, but I ultimately decided against them and for the phrase *for the future* because by phrasing it in this general way we are not limiting interpretations to *a home is money* as involving the monetary value of a home or house as real estate, but instead we allow for a metaphoric reading of *money* as implying that a home is valuable, for instance, in the sense that it brings happiness for the whole family for the future. Also, having *investment* in the continuation, in addition to the word *money* being the explicit metaphor source concept seems redundant and could thus result in lower plausibility ratings than when we used *for the future* because it would add no new information that could aid inferential processing. Also note that the continuation *for the future* includes the sense of investment as well as the sense labelled **personal** in the previous chapter.

- (40) *Ein Freund ist ein Magnet (für Freude).*
 A friend is a magnet (for happiness).
 ‘A friend is a magnet (for happiness).’

The continuation *für Freude* (*for happiness*) is, by a wide margin, the most common continuation created by both English and German monolinguals. More specifically, *Freude* was the

word that German speakers used most often in continuations to this metaphor and *happiness* was the word that English speakers used most often. *Freude* could also be translated as *joy* or *fun*, another common word used in the continuations by English monolinguals, but notice that *fun* is semantically narrower than *happiness*. German monolinguals also used the word *Glück* for *happiness* in continuations alluding to the same sense, but I decided against *Glück* in the German translation because it could also be taken to mean *to be lucky* by German speakers. In the end I chose *Freude* for the German translation of the context because it was semantically less ambiguous than *Glück* and also more common in the continuations created by German monolinguals, and I decided on *happiness* for the English translation because it was semantically more general than *fun*, matched the meaning of *Freude* as well as *joy*, but was more frequent than both *fun* or *joy* in the continuations created by English monolinguals. In other metaphors, whenever German monolinguals used the word *Glück* (*happiness*), English monolinguals also used the word *happiness* to allude to the same sense. So although *happiness* and *joy* might be very close in meaning, it seems monolinguals preferred *happiness* in all contexts where German speakers choose *Glück*.

Contexts created for the rest of the metaphors

Cross-linguistically shared metaphors

- (41) *Das Leben ist eine Reise (auf der Straße der Ungewissheit).*
 The life is a journey (on the road of the uncertainty).
 ‘Life is a journey (on the road of uncertainty).’

The closest English cognate to German *Straße* is *street*, but *street* has undesirable connotations in English in the context of this metaphor while *road* comes closer to the meaning that *Straße* has in this context: when *street* and *road* are used to refer to the concept of a carriageway, German speakers would use the word *Straße* for both types of carriageways. *Straße* also occurs in idioms in German where the pathway is metaphoric, where it is a transition or change or transformation of some sort, as in *auf der Straße zum Glück* (*on the street to happiness*) or *auf der Erfolgsstraße* (*on the street of success*). In English, *street* can also be used metaphorically, as in *you can get a good bargain on the high street*. However, notice that *street* as part of the compound noun *high street* is a metonym for chain stores of whatever level of prestige (and lately, physical as opposed to online shops) and this metonymic sense is distinctly not the metaphoric path implied by *Straße* in the previous example in German. *Road*, however, occurs in a number of idioms in English where it is a metaphoric path that represents a change, a motion, a progression towards a goal, as in *road to perdition*, *the road to freedom*, or *to hit the road*. In the metaphor in (41), the *journey* is also a metaphoric journey. Hence, in order for the continuation to allude to this metaphoric journey, it must conceptually incorporate the notion of a metaphoric path and *road* is more likely to do so for English monolinguals than *street*. Therefore, *road* is a better translation for *Straße* because they both can imply a path in the same metaphoric way and their idiomatic use is very similar. Alternatively we could use *way* in English and *Weg* (*way*) in German, which can also refer to a metaphoric path in both languages in the same way that *road* and *Straße* do: for instance, *on the way to fame and success* can be translated into German quite literally as *auf dem Weg zu Ruhm und Erfolg*, using the word *Weg* in the same sense as *way* in English. English and German share other idioms where *way* and *Weg* are used in the same sense and context, as in *Tom stood in the way of Linda’s promotion*

(*Tom stand Lindas Beförderung im Weg*). Another idiom, this time one where *Weg* occurs in German, is *auf dem Holzweg sein*. The closest in meaning might be the English idiom *to be on the wrong trail*, but literally translated the idiom would be *to be on the wooden way*. Although *way* and *Weg* can be used to imply a metaphoric path rather than an actual path, notice that there is an element of definiteness to them that is not present in *road* and *Straße*: if we used *way* instead of *road* in the continuation we would get *on the way to uncertainty* which would imply that whoever is on that journey is bound to end up in uncertainty. Using *Weg* in the German continuation would yield the same implicature. So when we use *way*, the continuation seems to imply that uncertainty is the goal of the journey, whereas when we use *road* and *Straße* it implies that there will be uncertainty along the way. I therefore think it is better to use *road* and *Straße* than *way* and *Weg*. *Ungewissheit*, which literally translates to *uncertainty*, implies that there might be ups and downs and unforeseeable obstacles during the journey. It would seem this makes for a better continuation than one that takes a stance on whether the journey will be pleasant or not and by implying that the journey might include parts where it is unpleasant and pleasant in succession, we make a very general statement that must surely hold true for most journeys. Also, we commonly find *life* and *uncertainty* paired up in expressions such as *life is full of uncertainty* or *there is no uncertainty in life* and this co-occurrence is true for English and German. It would thus seem that English and German participants should not be surprised to see *Ungewissheit* and *uncertainty* in a continuation to the metaphor *life is a journey*.

- (42) *Zeit ist Geld (in der Arbeitswelt).*
 Time is money (in the working world).
 ‘Time is money (in the working world).’

Arbeits-welt literally translates to *working world* which has the same general economic sense in both English and German and it would naturally be activated by both English and German speakers within the reference frame activated by *Geld (money)* and the meaning of the metaphor as a whole. It is therefore a natural candidate for the context that should make the implied metaphoric analogy between the target concept *Zeit (time)* and the source concept *Geld (money)* more feasible. Since the idiomatic expression *time is money* is a common motto associated with the working world, it is also quite natural to suggest that that motto holds true *in the working world*.

- (43) *Sehen ist Glauben (in seiner Verlässlichkeit).*
 Seeing is believing (in its reliability).
 ‘Seeing is believing (in its reliability).’

Verlässlichkeit is the literal translation of *reliability*. Since the implied metaphoric analogy between the source concept *Glauben (believing)* and the target concept *Sehen (seeing)* hinges on the durability of belief, the assumption that that belief should also be reliable and dependable should be a natural inference to make for both German and English speakers. Note that the anaphoric pronoun *seiner* in the German translation morphologically agrees with and refers to *Glauben (believing)* because *Glauben* is masculine. The form of the pronoun is morphologically ambiguous because it also agrees with *Sehen (seeing)*, which is neuter, and the form of the pronoun would also be *seiner*, but semantically the antecedent is *Glauben*. All other things being equal, this morphological ambiguity should work in favour of the plausibility of the

metaphor because if the pronoun can refer to both the source concept *Glauben* (*believing*) and the target concept *Sehen* (*seeing*) then morphological agreement with both should work to the advantage of the implied metaphoric analogy between source and target concept.

- (44) *Liebe ist blind (am Anfang).*
Love is blind (in the beginning).
'Love is blind (in the beginning).'

Anfang is a literal translation of *beginning*. In both English and German, we have a choice between the preposition *at* and *in*: *at the beginning* tends to occur in contexts where it has a more concrete sense as in *at the beginning of the line*, whereas *in the beginning* implies a more metaphoric use of *beginning*, as in *in the beginning, there was light* or *in the beginning of the relationship*. In German we find a similar distinction between the two prepositions, but note that the use of each preposition does not map one-to-one onto the use of the two English prepositions: *am Anfang* can either be used in context where English speakers would use *at the beginning* as in *am Anfang der Schlange (at the beginning of the line)*, but it can also be used in context where English speakers would use *in the beginning*, as in *am Anfang der Beziehung (in the beginning of the relationship)*. Note that for German speakers, *am Anfang* in *am Anfang der Beziehung* implies the same metaphoric sense of *Anfang* (*beginning*) as *in the beginning* implies for speakers of English in *in the beginning of the relationship*. In German translations of the Bible we find *im Anfang*, which although considered archaic and antiquated, is usually preferred because it is closer to the Hebrew original. For this reason *im Anfang* is mostly avoided outside this archaic biblical sense. I therefore chose *am Anfang* for the German translation because it is more natural, without the biblical, archaic connotations, and can be used metaphorically and *in the beginning* in English because in English the use of the preposition *in* signals the metaphoric sense of *beginning*. In order for the continuation to be compatible with the implied metaphoric analogy between the source concept *blind* and the target concept *Liebe* (*love*), we need to invoke *Anfang* (*beginning*) in a metaphoric rather than concrete sense, which is precisely what the choice of prepositions in the English and German translation should give us.

- (45) *Rache ist süß (mit Genugtuung).*
Revenge is sweet (with satisfaction).
'Revenge is sweet (with satisfaction).'

The metaphoric meaning hinges on the positive reward of revenge. *Satisfaction* seemed like a natural choice of words that would allude to this type of reward. I translated *satisfaction* as *Genugtuung* in the German continuation. Another translation for *satisfaction* would be *Befriedigung*, but *Befriedigung* has strong sexual connotations in German, which are undesirable when we wish to make the continuation compatible with the metaphoric analogy between the source concept *süß* (*sweet*) and the target concept *Rache* (*revenge*). *Genugtuung* is the natural choice of words in the context of *Rache* (*revenge*). Together these two reasons strongly favour *Genugtuung* over *Befriedigung* as the translation of *satisfaction*.

- (46) *Geduld ist eine Tugend (in allen Lebenslagen).*
Patience is a virtue (in all circumstances of life).
'Patience is a virtue (in all areas of life).'

The metaphor has strong moral undertones and so a continuation must apply to the same domain as morality, which is life as a whole. The compound noun *Lebenslagen*, which I translated

as *areas of life*, literally translates to *life levels* or *life circumstances*. It is a common idiomatic expression used in this context which should make it highly familiar and natural to German speakers. Similarly, the phrase *in all areas of life* occurs quite commonly with this choice of words, which makes it a good candidate for a translation. Ultimately, *Lebenslagen* and *areas of life* conceptually imply the same thing: in both English and German it means not only the totality of a person's personal life and life history, but also cross-cuts across all strata of society when we take life to refer not just to the whole of one person's life, but the lives of everyone. Both the English and German phrasing allow these myriad interpretations.

English-specific metaphors

(47) *Reden ist billig (in seiner Überschwänglichkeit).*

Talk is cheap (in its exuberance).

‘Talk is cheap (in its exuberance).’

The proverb *talk is cheap* implies that it is cheap because it is free, but also unbounded. The implied metaphoric analogy between the source concept *cheap* and the target *talk* hinges on the double meaning of *cheap* as either *inexpensive* or *of low quality*. The adjective *billig* has the same two senses and therefore is a good literal translation. The unboundedness of talk is what makes it cheap in the second sense and so the continuation must allow for this sense rather than block it. That talk can be inappropriate when it is unbounded is also evidenced in the cross-linguistically shared metaphoric proverb *Reden ist Silber (und/aber) Schweigen ist Gold* (*speech is silver (and/but) silence is golden*). The unboundedness feature of talk should thus be highly salient to both English and German speakers in the context of the metaphor *talk is cheap* as well. *Exuberance* seemed like a good choice of word to reinforce the unboundedness feature thereby making the continuation compatible with the metaphor. *Überschwänglichkeit* is a literal translation of *exuberance* with all of the same connotations in English as in German, especially when it comes to poetic contexts.

(48) *Worte sind Dolche (aus deinem Munde).*

Words are daggers (out of your mouth).

‘Words are daggers (out of your mouth).’

The metaphor is reminiscent of another cross-linguistically shared proverb: *the pen is mightier than the sword* (*die Feder ist mächtiger als das Schwert*). *Pen* is to *words* as *words* is to *language* and *daggers* and *swords* both metaphorically stand for fighting or any kind of struggle. If a struggle can be either physical or verbal (emotional and psychological) then words can hurt like a dagger or sword wound. The continuation uses the noun *mouth* because it is also commonly associated with language and words and would be common in the reference frame associated with a verbal struggle and German and English speakers are likely to use the same reference frame in the context of this metaphor. Note the use of the possessive pronouns *your* and its German equivalent *deinem*, which is quite common with body parts and should therefore be quite natural in the context of *Munde* (*mouth*). Lastly, note that it is *Munde* not *Mund* in the German translation of the continuation. Both *Munde* and *Mund* mean *mouth*, but *Munde* would sound more poetic to a German speaker than *Mund*, which is the common Modern High German form, because *Munde* is the more old-fashioned form. In all other metaphors I have usually avoided choosing words that were particularly poetic, but it seemed the more

appropriate choice of the two here because the metaphor *words are daggers* sounds very poetic to begin with, especially given its similarity to the equally poetic *the pen is mightier than the sword*. And notice that it is not just the choice of words that makes it *sound* poetic, but it is the subject matter itself that is poetic.

- (49) *Die Liebe ist ein Schlachtfeld (der Nebenbuhler).*
 The love is a battlefield (of the love-rivals).
 ‘Love is a battlefield (of love-rivals).’

Liebe can optionally have a definite determiner in German or not. When it occurs with a definite determiner it invokes a more general, poetic sense. The effect of the determiner is the reverse its presence has in English, e.g. *love is all around* (love in general) versus *the love I feel for you* (a specific type of love or romantic relationship). *Nebenbuhler* is the literal equivalent of *love-rivals* with all its senses and *Nebenbuhler (love-rivals)* is part of the reference frame that both English and German speakers would activate upon reading the metaphor. The continuation should therefore be compatible with the metaphor.

- (50) *Rache ist eine Zicke (in ihrer Niederträchtigkeit).*
 Revenge is a bitch (in its baseness).
 ‘Revenge is a bitch (in its baseness).’

The profanity *bitch* is the English equivalent of the German profanity *Zicke (young goat)* in all contexts. The profanity is central to the idiomatic meaning of the English-specific metaphor. German speakers agree with the English speakers that revenge can haunt and bring down the one seeking revenge. Hence, if German speakers can transfer the profanity of *Zicke* from its usual context to the context of the metaphor, then they should be able to access those aspects of their concept of revenge that English speakers would also access as part of the idiomatic meaning in English. *Niederträchtigkeit*, a literal translation of *baseness*, would then be a highly salient feature of the concept *revenge* for both English and German monolinguals which would make the continuation conceptually plausible and compatible with the metaphor.

- (51) *Ein Zuhause ist eine Burg (in die man sich zurückziehen kann).*
 A home is a castle (in(to) which one oneself retreat-to can).
 ‘A home is a castle (to retreat to).’

The verbs *to retreat to* and *sich zurückziehen* have the same meaning and use: they can refer to a military retreat or seeking refuge in a strategic sense and they can be used in a casual non-military context, as in *shall we retreat to the lounge*, in English, or *sich in seine eigenen vier Wände zurückziehen (to retreat home, literally, to retreat to within your own four walls, a common idiom)*, in German. The fact that it can be used in the military and casual sense in both languages is what makes it a prime candidate to occur in the continuation because this way it can both refer to the metaphor source concept *Burg (castle)* in its military-strategic sense and to the target concept *Zuhause (home)* in its casual sense. Due to the two contextual uses, it should also be part of the reference frames that English monolinguals would activate for the idiomatic meaning of the metaphor and the reference frames of *Zuhause (home)* and *Burg (castle)* that German monolinguals should activate upon reading the source and target concept. The continuation should therefore be conceptually plausible to both English and German speakers.

- (52) *Lob ist kein Pudding (wenn es hohl ist).*
 Praise is no pudding (if it hollow is).
 ‘Praise is not pudding (if it is hollow).’

The continuation is not a prepositional phrase as the others but should be conceptually plausible for both English and German speakers. German speakers are certainly aware that praise can be hollow and empty and if they can infer the implied metaphoric analogy between *Lob* (*praise*) and *pudding*, they should be able to make the conceptual connection to hollow praise. The continuation makes the hollowness of praise salient through two devices: (1) the use of the adjective *hohl*, a literal translation of *hollow*, and (2) the use of the anaphoric pronoun *es*. In the German translation, the anaphoric pronoun *es* refers back to *Lob* (*praise*, which is neuter in German). In German, anaphoric pronouns agree with the noun they refer to and this agreement is morphologically marked. The referent of the pronoun is unambiguous because if it referred to *Pudding* (which is masculine) its form would have to be *er*. Therefore, if German speakers, upon reading *Lob* (*praise*), activate the feature of hollow praise as part of its reference frame then they should find the continuation conceptually plausible and compatible with the metaphor in the same way that English monolinguals would.

German-specific metaphors

Two of the continuations for German metaphors were discussed above. Here are four further examples:

- (53) *Träume sind Schäume (auf dem Ozean der Wünsche).*
 Dreams are froth (on the ocean of wishes).
 ‘Dreams are froth (on the ocean of wishes).’

Recall that in the diachronic analysis of the corpus frequency of this metaphor we saw that in German it has, since Freud's mention of it in his 1911 book, been an established proverb and its frequency has been steadily increasing. Also recall that I pointed out that the reason why I opted for *froth* rather than *foam* as a translation of *Schäume* is because *foam* can refer to non-natural things, e.g. the kind of foam used in construction work, while *froth* is necessarily natural. A second reason was that *froth* is the preferred choice of words used in translations of Freud's work. A third reason is that *froth* can have a more poetic sense than *foam*, which works to the advantage of the implied metaphoric analogy between the source concept *Schäume* (*froth*) and the target concept *Träume* (*dreams*). English speakers should be likely to activate both *froth* and *foam* as part of the reference frame of *ocean* that occurs in the continuation. Hence, *ocean* connects the continuation back to *froth* in the metaphor. Similarly, *Wünsche* (*wishes*) in the continuation connects back to *dreams* as dreams are often an expression of hopes, wishes, and desires. *Wishes* should therefore be part of the reference frame that English speakers activate upon reading *dreams* and when they read *ocean* they should form the appropriate mental image of *froth* and *foam* of the surf pounding against rocks, an image that would be highly salient in the context of *ocean*. Therefore, English speakers should have all that is necessary to conjure up the same mental images that German speakers would experience upon reading the continuation in combination with the metaphor.

- (54) *Kunst ist eine Wissenschaft (des Experimentierens).*
 Art is a science (of the experimenting).
 ‘Art is a science (of experimenting).’

Recall that in the original discussion of the experimental material I discussed at length which meaning German and English non-academic speakers would assign to *Wissenschaft* (*science*) and I argued that they should be able to take it in the same broad sense of science as the academic effort that yields knowledge as opposed to a mere collection of facts or information, rather than the narrow academic sense of science as the natural sciences. And I demonstrated that even among academics, English and German universities will apply the label *science* to the humanities as well, i.e. also use *science* in the broad sense. And recall that the idiomatic meaning of the German-specific metaphor hinges on the morphology of the noun *Wissenschaft* (*science*) because it is made up of *Wissen* (*knowledge*) and *-schaft*, which refers to any craft or discipline that creates that which it denotes. Hence, a literal translation of *Wissen-schaft* would be *a craft* or *discipline that creates knowledge*. That scientists conduct experiments is widely known, both within the English and German speaking world (even if non-academics might have a false belief about the exact nature of a scientific experiment). Hence, alluding to it in the continuation through the use of *Experimentierens* (*experimenting*) should reinforce the broad sense of *science* that English and German speakers are likely to apply. English and German speakers alike should also be conceptually familiar with the idea that artists also experiment, albeit in a different way. At the same time, when academics in the English speaking world try to raise public awareness for what scientists do they may, for instance, refer to their work as *creative* or *imaginative* and might call a particular scientific contribution *inspired*, the same terminology English speakers would use to talk about artists. The use of the word *experimenting* in the continuation should therefore be conceptually plausible (1) for the source concept *science* and target concept *art* individually and (2) for the implied metaphoric analogy between the source and target concept collectively. English monolinguals should therefore have a better chance of inferring a meaning similar to the idiomatic meaning that German speakers expect when they read the continuation.

- (55) *Schönheit ist eine Eintrittskarte (zum Ruhm).*
 Beauty is a ticket (to the fame).
 ‘Beauty is a ticket (to fame).’

Schönheit (*Beauty*) is a feature that is commonly attributed to those who are famous. Thus, *Ruhm*, a literal translation of *fame*, seems like a natural continuation to the metaphor that should refer to a highly salient feature that English speakers might activate as part of the reference frames of *beauty* and of *fame*. If English speakers are able to conceptualise how *beauty* may not just be a common feature of celebrities, but often a prerequisite in order to become famous, then they should be able to infer an implied metaphoric analogy between the source concept *ticket* and the target concept *beauty* that allows them to understand a ticket as something that allows access or entry to something that is otherwise reserved in a way similar to how *beauty* might grant access to *fame*.

- (56) *Mitgehen ist Mitgefangen (in den Augen des Anklägers).*
 Hang-alongside is caught-alongside (in the eyes of the prosecution).
 ‘Following is suffering (in the eyes of the prosecution).’

Recall that when native speakers of German think of the idiomatic meaning of the metaphor, they do not realise that conceptually the order of *mitgehen* and *mitgefangen* is the reverse of the order in the sentence (linguistic form). Note, in German, *des Anklägers* is singular while

of *the prosecution* in English, though grammatically singular, could semantically also refer to multiple persons. Alternative to *Ankläger*, German also allows the use of *die Anklage* to refer to the prosecution, but since *die Anklage* can also refer to *the allegations* raised against the prosecuted, I opted for *Ankläger* instead because it can only refer to a person or group of people, not a thing. The use of *Ankläger* and *prosecution* should cause experimental participants to activate the reference frame of court cases, the frame within which the idiomatic meaning of the German metaphor is to be understood. The idiom *in den Augen der* has exactly the same meaning and contextual use as *in the eyes of* in English. Therefore, English speakers should have a good chance to infer a meaning for the metaphor that is similar to the one that German speakers would have as an idiomatic meaning.

Novel metaphors

- (57) *Ein Zuhause ist ein Schwamm (des Wohlbehagens).*
 A home is a sponge (of feeling comfortable).
 ‘A home is a sponge (of comfort).’

Recall that for the metaphor *a home is money* monolinguals and bilinguals commonly associated *home* with *comfort*, hence this seemed like a context that would make the quite nonsensical novel metaphor *a home is a sponge* conceptually plausible. The German noun *Wohlbehagen*, which I have translated as *comfort*, is a compound made up of *wohl* (*well* as in *well-being*) and *behagen*, which means *feeling* or *experiencing* and also commonly occurs in other compounds such as *Unbehagen* (*feeling uncomfortable*), the antonym of *Wohlbehagen*. Note that the juxtaposition of the concept of *Wohlbehagen* (*the feeling of being comfortable*) with the noun *Schwamm* (*Sponge*) would be quite natural for a German speaker while it would be unnatural for an English speaker. I have therefore decided to translate it as *comfort*, the state of being comfortable, rather than the cumbersome phrase *the feeling of being comfortable*. Note that this is a marginal conceptual difference between a state and a feeling or experience: the state that the English wording alludes to is the end state of the feeling and experience that a person undergoes in the German wording of the continuation. If the continuation reinforces the activation of the highly salient feature of a home being comfortable, then both English and German speakers should be able to infer how a *home* can soak up and preserve *comfort* in the same way that a *sponge* can soak up and preserve water.

- (58) *Schönheit ist ein Computer (der Berechenbarkeit).*
 Beauty is a computer (of calculability).
 ‘Beauty is a computer (of calculability).’

Computer should not be part of the reference frame of *Schönheit* (*beauty*) and it would be quite hard to see what the implied analogy between the metaphor source concept *computer* and the target concept *beauty* is thus supposed to be, but at least the word *Berechenbarkeit* (*calculability*) in the continuation is a salient concept in the reference frame of *computer*. English and German speakers might be quite at a loss here. Nevertheless, whatever they are able to take away from the implied metaphoric analogy, the continuation should at least be conceptually plausible with reference to the source concept *computer*. *Berechenbarkeit* is a literal translation of *calculability*, the same term that would be used in German mathematics and computer science.

- (59) *Schönheit ist Geduld (auf dem Weg zur Perfektion).*
 Beauty is patience (on the way to the perfection).
 ‘Beauty is patience (on the way to perfection).’

Geduld (*patience*) might already be part of the reference frame that English and German speakers activate upon reading *Schönheit* (*beauty*), which might make the implied metaphoric analogy between the source concept *Geduld* (*patience*) and the target concept *Schönheit* (*beauty*) already conceptually plausible. The notion of *Perfektion* (*perfection*) would probably also be part of the same reference frame and notice that would also be in the reference frame of *Geduld* (*patience*) because *perfection* often requires *patience*. It would thus reinforce the connection between the two reference frames and thus the implied metaphoric mapping. I thought it would have sounded less natural to have *Schönheit ist Geduld zur Perfektion* in German and less natural to have *beauty is patience to perfection* in English because in both languages this juxtaposition seems to imply that *patience* is to be perfected rather than the implicature that *beauty* requires *patience* in order to achieve perfect beauty, the intended metaphoric meaning. I therefore added *auf dem Weg* which literally translates to *on the way* and recall that I pointed out earlier for context (41) that the idiomatic use of German *Weg* is similar to that of English *way*.

- (60) *Liebe ist Mord (für Selbstsüchtigkeit).*
 Love is murder (for selfishness).
 ‘Love is murder (for selfishness).’

The metaphor *love is murder* has a straightforward interpretation, where *murder* is taken to mean “acutely difficult or problematic,” an *ad hoc* concept compatible with the target concept LOVE. The metaphor *love is murder* does not appear in the corpora analysed and at the time the experiments took place, only one, less-known novel with that title had appeared in English. It therefore seemed reasonable at the time to classify it as a novel metaphor. Since then, it appears in several English works of fiction. It should therefore be reclassified as English-specific if the experiments were to be reproduced. Notice how *Selbstsüchtigkeit* (*selfishness*) in the continuations can equally apply to *love* and *murder*: *selfishness* is not a desirable attribute of a person in a romantic relationship. In search for possible analogies between *love* and *murder*, it is highly salient that we are looking for negative attributes such as *selfishness*. *Selfishness* might also be the kind of negative trait in a person’s character that English and German speakers would activate as part of the reference frame that contains common motives for *murder*. In this twofold way, *selfishness* should be conceptually plausible and compatible with the metaphor because it refers to salient features of the source and target concept as well as being a possible connection between them that could serve as the metaphoric analogy between the two concepts.

- (61) *Liebe ist eine Burg (aus dem Märchen).*
 Love is a castle (out of the fairy tale).
 ‘Love is a castle (of fairy tales).’

Note that *Märchen* in German is singular while *fairy tales* in the English translation is plural. We want the continuation to talk about *fairy tales* in general and in order for the continuation to imply this general sense, German requires that we use singular while English requires that we use plural. For instance, if we changed the English version to singular, *love is a castle of a fairy tale*, it seems as though we were talking about a specific fairy tale in which a particular

castle has some symbolic function for a couple's love. This is not what we want in order to test the conceptual plausibility of the implied metaphoric analogy between the general concept of *love* and the general concept of *castle*. Also note that the preposition used in German is *aus*, *aus dem Märchen*, which literally translates to *out of the fairy tale*. A literal translation of the English possessive construction *of fairy tales* would be to use the genitive, *des Märchens* (*of the fairy tale*, the definite determiner is morphologically marked as genitive, *des*, and so is the noun by use of the suffix *-s*). However, I decided against the possessive noun phrase in the German version because the phrase *aus dem Märchen* is a very common idiomatic expression used in the context of fairy tales. It is also often used idiomatically in comparisons: *wie aus dem Märchen* which literally means *as if out of a fairy tale*, but which can more generally be taken to mean *picturesque* with a fantastical element to it.

- (62) *Zeit ist Glauben (mit Stetigkeit).*
 Time is believing (with continuity).
 'Time is believing (with continuity).'

Stetigkeit is a literal translation of *continuity*, which both in English and German means continuity in either a mathematical sense and in the general sense. *Zeit* (*time*) is continuous and continues to be in a similar way that *belief* can be continuous and with a personal effort can continue to be. The continuation thus applies to both the metaphor source concept *believing* and the target concept *time* individually and by being part of their individual reference frames it also reinforces a potential metaphoric analogy between them that could be a feasible interpretation of the metaphor.

Throughout the discussion of the various context continuations we have seen time and time again that even when monolinguals of the L2 would not know the idiomatic meaning that L1 monolinguals know, the continuations should still make the same features of the metaphor source and target concepts, of their individual reference frames, as well as the implied metaphoric analogy conceptually more plausible. I hope this overview was sufficient to convince you that for all intents and purposes, the context continuations should help improve the conceptual plausibility and, thus, the cross-linguistic intelligibility of all metaphor types (cross-linguistically shared, L1 language-specific, L2 language-specific, and novel metaphors). For a complete list of all contexts, see Table (A.1) in Appendix A. In the next section we look at how these continuations were incorporated into the experimental design.

8.4.4 Procedure

In three experiments English monolinguals, German monolinguals, and English-German bilinguals saw metaphors in context, i.e. followed by continuations (mostly) in the form of prepositional phrases, e.g. *knowledge is power over those without it*. Eight of these contexts were created by native speakers in part 2 of the out-of-context experiments and in the previous section I described how I selected those out of the variety of continuations participants created that (1) in the case of cross-linguistically shared and language-specific metaphors, alluded to metaphors' conventional idiomatic meaning and (2) in the case of novel metaphors, were conceptually plausible by either alluding to features of the metaphor source and target concept or the implied metaphoric analogy that was conceptually most plausible. I also described how I created the rest of the contexts by following the same two premises. The experimental procedure was exactly the same as in part 1 of the out-of-context experiments: English monolinguals saw

the same English version and wording of the instructions used in the out-of-context experiments (see Chapter 7), German monolinguals and bilinguals saw the instructions in German using the same wording as in the out-of-context experiments. Experiments conducted in the lab used the E-Prime software. Metaphors were displayed one at a time. Experiments conducted via the internet showed ten metaphors per slide. In both experiments conducted in the lab and via the internet, the order of metaphors was randomised and counter-balanced so as to avoid having metaphors of the same type (e.g. novel–novel) following right after another and in experiments conducted via the internet the lists of ten metaphors were additionally counter-balanced such that there was an equal number of metaphor types per slide (in order to avoid, for instance, having, say, five out of the ten be cross-linguistically shared metaphors). Recall that I showed in Chapter 7 for the out-of-context condition that plausibility judgements collected in the lab and those gathered via the internet are qualitatively and quantitatively comparable. The only difference in the procedure in the in-context experiments is that participants rated metaphors on a 7-point Likert scale while in the out-of-context experiments it was a 5-point scale. I had changed this in the hope that plausibility judgements would be more fine-grained and thus more sensitive to the effects of context. However, as we will see from the z-scores, this turned out not to be the case. The ratings are, overall, as fine-grained as in the out-of-context experiments. Hence, results of the in-context experiments should be comparable to those of the out-of-context experiments and independence of samples is guaranteed by the fact that new (different) monolinguals and bilinguals were recruited for the in-context experiments.

In anticipation of an important issue of this chapter, I want to highlight that, in the experiments, *cross-linguistic metaphor intelligibility* is defined in two ways: (1) An L2 metaphor is *weakly intelligible* if participants judge it to be conceptually more plausible than novel metaphors, and (2) it is *strongly intelligible* if it is judged to be conceptually as plausible as cross-linguistically shared metaphors. Both definitions of cross-linguistic intelligibility define it in relative terms: *weak intelligibility* is intelligibility relative to novel metaphors and *strong intelligibility* is relative to shared metaphors. In other words, the plausibility ratings of novel and cross-linguistically shared metaphors are the standard, the control conditions, against which we evaluate L2 metaphor plausibility. In the next section, I will discuss how context has affected L2 metaphor intelligibility. Keep in mind that calling the plausibility ratings of novel and shared metaphors control conditions assumes that their ratings remain constant. However, context affects their ratings which means they are not ideal control conditions. The plausibility ratings of cross-linguistically shared metaphors are higher than the ratings of any other metaphor type in all of the out-of-context and in-context experiments. The ratings of the novel metaphors are the lowest ratings in all of the experiments. I will therefore argue that although context affects the perceived plausibility of shared and novel metaphors, because they have the highest and lowest ratings in all experiments, we can use them as standards relative to which we can evaluate cross-linguistic intelligibility, but, importantly, we need to take the effect that context has on their ratings into account when we evaluate the statistical significance of differences between their ratings and those of the L2 metaphors. I will argue that in order to take the effect of context into consideration, we need to compare the rate of change in plausibility ratings from out-of-context to in-context experiments. From this approach we not only benefit by validating the operationalisation of cross-linguistic metaphor intelligibility in its weak and strong sense, but we also see the systematic effect that context has on metaphors: (1) as we would expect, context *increases* the plausibility of metaphors for which speakers

have no conventional idiomatic meanings or conventional expectations of contextual relevance if that context is conceptually compatible with the metaphors, and (2) context *decreases* the perceived plausibility of metaphors for which speakers have conventional idiomatic meanings because these conventional meanings constrain which contexts speakers expect.

8.4.5 Design

The outcome variable is plausibility judgements rated on a 7-point Likert scale. Predictor variables are (1) metaphor type (English-specific metaphors, German-specific metaphors, cross-linguistically shared metaphors that occur in both English and German, and novel metaphors that occur neither in English nor German), (2) speaker group (English native speakers who do not speak German = English “monolinguals,” German native speakers who do not speak English = German “monolinguals,” and English-German bilingual speakers who are born native speakers of German, are fluent in English, and reside in an English-speaking country), (3) age, (4) gender, and (5) metaphors’ corpus frequency. We would expect that some participants tend to avoid giving the highest or lowest plausibility ratings. At the same time we would expect that metaphors of the same type will uniformly receive similar ratings. Since participants’ use of the rating scale and the variation in plausibility ratings given to metaphors of the same type is not independent of each other, linear mixed effects models are used in all analyses.

8.5 Results

In order to answer question 2 we are technically only interested in three things: (1) has the presence of context changed monolinguals’ plausibility judgements of L2 metaphors and if so, has their plausibility increased or decreased? Recall that contexts for eight of the metaphors were created by native speakers in the first set of experiments, the rest were created by me, all of which were such that they should increase metaphor plausibility. (2) Has context had an effect on bilinguals’ plausibility ratings and have they increased, as we would expect, or decreased? (3) Has context had the same effect on monolinguals and bilinguals? As we will see shortly, context has also had an effect on monolinguals’ perception of the plausibility of L1 metaphors and I will argue that (1) this contradicts the predictions made by CMT, and (2a) it prompts us to take a broader view on the effect that speakers’ familiarity with conventional meanings has on how they perceive and judge conceptual plausibility and (2b) since bilinguals are familiar with more conventional meanings for the language-specific metaphors than monolinguals, they incorporate contextual information differently for comprehending these metaphors. Before we get to this, let’s look at the results of the statistical analyses.

All statistical tests are linear mixed effects models that use Formula (8.1) as a model:

$$y = X\beta + Z_1b_1 + Z_2b_2 + \varepsilon \quad (8.1)$$

where y are the plausibility judgements, β is the fixed effect of metaphor type, b_1 is the random effect of by-subject variation in plausibility ratings, b_2 is the random effect of by-item variation in plausibility. X is a matrix of random intercepts (baseline plausibility ratings per metaphor) that regress metaphor type onto variation in plausibility judgements. Z_1 and Z_2 are matrices containing random intercepts *and* random slopes that relate the random effects to the variation in plausibility judgements y . In other words, the full model tests how well *metaphor type* can

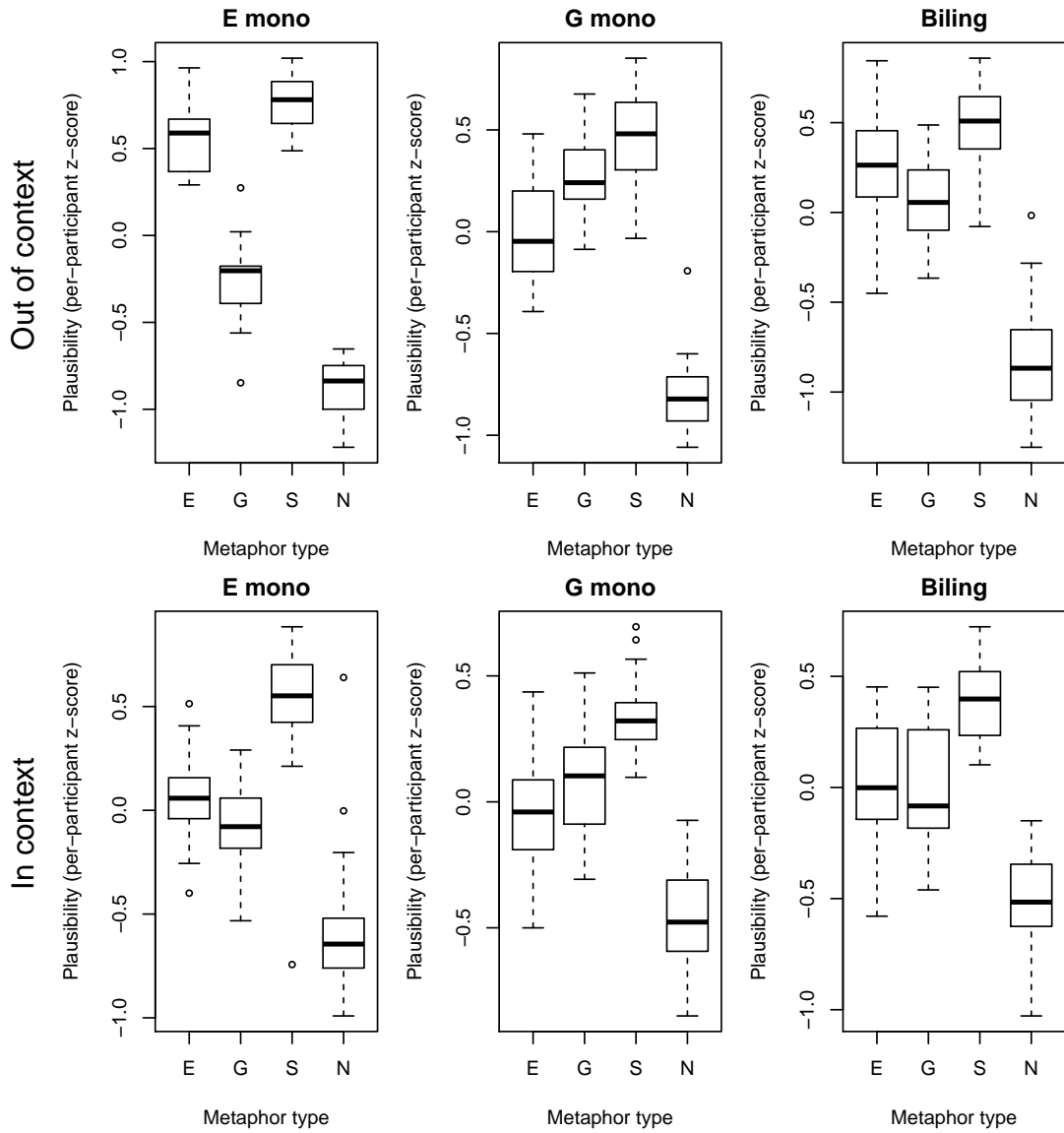


Figure 8.1: The y-axis shows median plausibility ratings per metaphor type, where ratings are converted to per-participant z-scores. Boxes and whiskers indicate between-participant variation in z-scores. The top row shows the out-of-context condition, originally presented in Figure (7.3); the bottom row shows the in-context condition.

Table 8.1: Variance explained by each fixed and random effect in the linear mixed effects models: in-context plausibility judgement models. Each value is a ratio of the variance accounted for by that effect and the total model-wide variance. These ratios are effectively R^2 estimates ranging from 0 to 1 (for ease of comparison, given here as percentages). The higher the value, the more of the overall variance is explained by the respective effect.

Effect	E mono	G mono	Biling
Metaphor type	57.6%	44.3%	52.9%
By-subject	6.0%	11.5%	11.3%
By-item	24.1%	23.4%	22.2%
Error term	12.3%	20.8%	13.6%

predict variation in plausibility ratings y by comparing the amount of variation in plausibility that is accounted for by by-subject and by-item variation. Recall that the main reason for using linear mixed effects models as opposed to linear fixed-effects-only models and ANOVA in the previous chapter was to account for by-subject and by-item variation simultaneously in the analysis, which would have been violated in ANOVA. In order to make the mixed models more conservative, i.e. more robust against Type I errors, random intercepts and slopes were fitted for all experimental participants and items as recommended by Schielzeth & Forstmeier (2009) and Barr et al. (2013). R (R Core Team, 2013) and the lme4 package (Bates et al., 2014) were used to perform the linear mixed effects analyses. The linear mixed effects models of this chapter are thus directly comparable to the models in the previous chapter because the model formulas and statistical analysis are identical.

Figure (8.1) shows the results of the in-context experiments in Panel B. In order to look for changes in plausibility judgements from the out-of-context condition to the in-context condition, I have repeated the z-score graphs from the out-of-context experiments in Panel A of Figure (8.1).

Recall that the vector of random error terms ε of the model formula (8.1) is not actually part of the fitted model, but is given implicitly by the residual of the fitted model, i.e. the amount of variation in plausibility judgements that the fitted model cannot account for through the fixed and the two random effects. The proportion of the variance in plausibility judgements y “explained” by the fixed and random effects to the “unexplained” variance ε is thus a measure of how well a mixed model fits the model. The effect size R^2 estimates given in Table (8.1) follow the method by Xu (2003), which compared to other methods I tried gave the best estimate of total variance. The model fitted to the plausibility ratings of English monolinguals accounts for 87.7% ($57.6\% + 6\% + 24.1\%$) of the overall variation in plausibility. The mixed model fitted to the ratings of German monolinguals accounts for 79.2% ($44.3\% + 11.5\% + 23.4\%$) of total variation. The model fitted to the judgements of English-German bilinguals accounts for 86.4% ($52.9\% + 11.3\% + 22.2\%$) of total variance in plausibility judgements. In the models of the ratings of English monolinguals and bilinguals, *metaphor type* accounts for most of the total variance and more than all of the random effects, including the random error term, taken together ($57.6\% > 6.0\% + 24.1\% + 12.3\%$ and $52.9\% > 11.3\% + 22.2\% + 13.6\%$). In all three models, the effect size of *metaphor type* is greater than either by-subject ($57.6\% > 6.0\%$, $44.3\% > 11.5\%$, and

52.9% > 11.3%) or by-item variation (57.6% > 24.1%, 44.3% > 23.4%, and 52.9% > 22.2%) and greater than by-subject and by-item variation taken together (57.6 > 6.0% + 24.1%, 44.3% > 11.5% + 23.4%, and 52.9% > 11.3% + 22.2%). Therefore, most of the variation in plausibility judgements that the three models fit so well is due to the fixed effect, *metaphor type*, and not due to by-subject and by-item variation.

Null models were constructed from the full models using Formula (8.2):

$$y_0 = Z_1b_1 + Z_2b_2 + \varepsilon \quad (8.2)$$

The null models are the full models *without* the variable whose effect on plausibility judgements we are interested in, *metaphor type*. Wilks theorem (1938) states that, given a sufficiently large number of samples, the likelihood ratio between two nested models is asymptotically χ^2 -distributed. Because the null models are derived from the full models, they are less complex and therefore nested within the full models. The χ^2 values that the likelihood ratios of comparing the full to the null models approximate can then be used to find the corresponding statistical significance. Note that in our particular case, the null models contain only random and no fixed effects. They are therefore directly notionally comparable to the null model assumptions of ANOVA. Our null models are “unaware” of a metaphor’s type; they are therefore models of the variance in plausibility judgements that we would normally expect, i.e. when metaphor type plays no role. We are thus taking “normally expected plausibility judgements” from the judgements we actually observed, instead of assuming that a Gaussian normal distribution would be a good null model. As we can see in Figure (A.7), plausibility judgements that disregard metaphor type scatter normally around the regression line of the fitted models. We thus know that our null models are “good” estimates of the normal distribution of plausibility judgements. Thus, comparing the full against the null models through χ^2 likelihood ratio tests is notionally comparable to asking whether the variation in plausibility judgements in a model that is “aware” of a metaphor’s type deviate significantly from what we would normally expect. The results of these χ^2 tests are shown in Table (8.2).

In order for a metaphor to be communicative, it must first be intelligible, and in order to be intelligible, the analogy implied between its source and target concept must be plausible. Therefore, a metaphor whose analogy is not plausible can also not be communicative. If CMT’s predictions are correct and metaphor comprehension is primarily a matter of conception, then our L2 language-specific metaphors, for which we know that their source and target concepts are cross-linguistically maximally similar in terms of their lexical meanings, Fregean senses, expectations of contextual relevance, and related mental imagery, should be as plausible to monolingual native speakers as L1-specific and cross-linguistically shared metaphors. If, however, RT and GS are correct, then we should expect to find that L2 metaphors are judged to be significantly less plausible than L1 and shared metaphors. Because participants see all metaphors with context intended to improve plausibility, however, CMT, RT, and GS would expect that L2 language-specific metaphors should have higher plausibility ratings than in the out-of-context condition.

English monolinguals find L2 metaphors less plausible and thus less intelligible than cross-linguistically shared metaphors (Table 8.2, $p = .012$). This is the same result we got in the out-of-context experiments, but note that the significance is less pronounced now than in the out-of-context condition. For German monolinguals, L2 metaphors are intelligible in both the

Table 8.2: In-context plausibility judgements on 7-point Likert scale. Best-fitting regression using mixed-effect modelling. Models fitted to responses of English monolingual native speakers (E mono, $R^2 = .877$), German monolingual native speakers (G mono, $R^2 = .792$), and English-German bilinguals (Biling, $R^2 = .864$). Intercept for metaphor type is cross-linguistically shared metaphors. P-values are likelihood ratios (χ^2 tests) between the linear mixed effects model that includes the variable of interest, *metaphor type*, and the model that does not (the null model). Significance codes: *** for $p \leq .001$, ** for $p \leq .01$, * for $p \leq .05$, † for $p \leq .1$.

Speaker	Predictor	B	St err	<i>t</i>	<i>p</i>
E mono	S (intercept)	5.6404	.2650	21.281	< .0001 ***
	E (L1)	-1.0500	.4521	-2.323	.025 *
	G (L2)	-1.2873	.4843	-2.658	.012 *
	N	-2.3989	.3796	-6.320	< .001 ***
G mono	S (intercept)	5.1429	.2949	17.438	< .001 ***
	E (L2)	-0.7220	.3853	-1.874	.063 †
	G (L1)	-0.5512	.3422	-1.611	.099 †
	N	-1.5137	.3385	-4.472	< .001 ***
Biling	S (intercept)	4.9592	.3483	14.239	< .001 ***
	E (L2)	-0.6933	.4562	-1.520	.098 †
	G (L1)	-0.7290	.4160	-1.752	.079 †
	N	-1.7727	.3782	-4.687	< .001 ***

strong (Table 8.2, $p = .063$, $p > .05$) and weak sense (Table 8.2, $p = .037$, $p < .05$). In the out-of-context experiments, they only found them intelligible in the weak sense (Table 7.5, $p = .002$, $p \ll .05$), but not in the strong sense (Table 7.5, $p = .022$, $p < .05$). It thus seems that context has not increased the conceptual plausibility of L2 metaphors in the eyes of the German monolinguals. Why context does not help monolinguals will have to be a topic for future research. At least for English speakers we can say that the fact that the cross-linguistic communicability of L2 language-specific metaphors increases with context tells us that the contexts that native speakers ‘spelled out’ is providing non-native speakers with information which they otherwise would not have. So this is further evidence that, without context, L2 language-specific metaphors are not cross-linguistically communicative.

In the in-context experiments, bilinguals judged both German L1 and English L2 metaphors to be intelligible in the weak sense (Table 7.5, English L2, $p = .023$, German L1, $p = .013$, both $p < .05$) and in the strong sense (Table 7.5, English L2, $p = .079$, German L1, $p = .098$, both $p > .05$). In the out-of-context experiments, bilinguals found English L2 metaphors to be intelligible in the strong sense (Table 7.5, German L1 and shared metaphors $p = .026$, $p < .05$, English L2 and shared metaphors $p = .303$, $p \gg .05$) and in the weak sense (Table 7.5, German L1 $p = .29$, $p \gg .05$, English L2 $p < .001$). For German L1 metaphors, their intuitions were reversed: they were not intelligible in the strong sense and also not in the weak sense. Hence, bilinguals’ ratings of the German metaphors have increased with context while ratings of the English metaphors have largely been unaffected. Nevertheless, English and German metaphors’ ratings are still very much the same in the in-context experiments ($B_E = -0.6933 \pm .4562$, $B_G = -0.7290 \pm .4160$, Table 8.2) as they were in the out-of-context experiments ($p = .43$, $p \gg .05$, Table 7.5).

The language specificity of idiomatic meanings and the effect of context

In the previous chapter, we saw that monolinguals' plausibility judgements of L2 metaphors are significantly lower than those of L1 and cross-linguistically shared metaphors. In this chapter, we see from Table (8.3) that monolinguals' L2 ratings increase with context:

$$|\Delta G.in - \Delta G.out| = |(0.84580 - 0.28695) - (0.84580 - 0.39887)| = 0.11192 \quad (8.3)$$

for English monolinguals and

$$|\Delta E.in - \Delta E.out| = |(0.69457 - 0.12443) - (0.69457 - 0.18142)| = 0.05699 \quad (8.4)$$

for German monolinguals. Before we try to understand the role of context, it is therefore important to ask "Are monolinguals generally more susceptible to language-specific idiomatic meanings than bilinguals?" In other words, we are interested in the interaction between speaker group and the language-specificity of metaphors' idiomatic meanings.

Monolinguals should be familiar with the idiomatic meanings of L1 and cross-linguistically shared metaphors. Bilinguals should be familiar with the idiomatic meanings of L1, L2, and shared metaphors. If speakers are susceptible to language-specific idiomatic meanings then they should rate L2 metaphors differently from L1 metaphors, while L1 metaphors should receive ratings similar to those of cross-linguistically shared metaphors. Let's start with the second part of the prediction. Monolinguals' odds of giving different ratings to L1 and cross-linguistically shared metaphors are

$$\frac{\exp[(Biling.G.out + Biling.E.out)/2]}{\exp[(Emono.E.out + Gmono.G.out)/2]} = \frac{\exp[(0.16267 + 0.09162)/2]}{\exp[(0.06019 + 0.06070)/2]} = 1.07 \quad (8.5)$$

times the odds of bilinguals. So there is hardly any difference between monolinguals' and bilinguals' treatment of L1 metaphors.

The first part of the prediction assumes that if speakers are susceptible to language-specific idiomatic meanings, their ratings of L1 and L2 metaphors should differ. In the out-of-context condition, this is only the case for monolingual but not bilingual speakers. English monolinguals' L1 and L2 ratings differ by

$$|\Delta G.out - \Delta E.out| = |(0.84580 - 0.39887) - (0.84580 - 0.06019)| = 0.33868; \quad (8.6)$$

German monolinguals' differ by

$$|\Delta G.out - \Delta E.out| = |(0.69457 - 0.06070) - (0.69457 - 0.18142)| = 0.12072; \quad (8.7)$$

more than bilinguals' difference in ratings of

$$|\Delta G.out - \Delta E.out| = |(0.71302 - 0.16267) - (0.71302 - 0.09162)| = 0.07105. \quad (8.8)$$

Table 8.3: Effect of context on plausibility judgements. Best-fitting regression using mixed-effect modelling. Models fitted to responses of English monolingual native speakers (E mono, $R^2 = .960$), German monolingual native speakers (G mono, $R^2 = .879$), and English-German bilinguals (Biling, $R^2 = .883$). Intercept for metaphor type is cross-linguistically shared metaphors rated out of context. P-values are likelihood ratios (χ^2 tests) between the linear mixed effects model that includes the variable of interest, *context*, and the model that does not (the null model). Significance codes: *** for $p \leq .001$, ** for $p \leq .01$, * for $p \leq .05$, † for $p \leq .1$.

Speaker	Predictor	B	St err	<i>t</i>	<i>p</i>
E mono	S.out (intercept)	0.84580	.03860	21.914	< .0001 ***
	S.in	-0.07239	.05816	-1.245	.0276 *
	E.out (L1)	-0.06019	.05825	-1.033	.200
	E.in (L1)	-0.24740	.07651	-3.233	< .0001 ***
	G.out (L2)	-0.39887	.08378	-4.761	< .0001 ***
	G.in (L2)	-0.28695	.08247	-3.479	< .0001 ***
	N.out	-0.64434	.06145	-10.486	< .0001 ***
	N.in	-0.47221	.06354	-7.432	< .0001 ***
G mono	S.out (intercept)	0.69457	.05905	11.762	< .0001 ***
	S.in	-0.00409	.08188	-0.050	.788
	E.out (L2)	-0.18142	.09727	-1.865	< .0001 ***
	E.in (L2)	-0.12443	.08941	-1.392	< .0001 ***
	G.out (L1)	-0.06070	.07095	-0.856	.083 †
	G.in (L1)	-0.09596	.07989	-1.201	.0015 **
	N.out	-0.48617	.08128	-5.982	< .0001 ***
	N.in	-0.25637	.08301	-3.088	< .0001 ***
Biling	S.out (intercept)	0.71302	.04827	14.771	< .0001 ***
	S.in	-0.04125	.07461	-0.553	.0362 *
	E.out (L2)	-0.09162	.07813	-1.173	< .0001 ***
	E.in (L2)	-0.13938	.08801	-1.584	< .0001 ***
	G.out (L1)	-0.16267	.06517	-2.496	< .0001 ***
	G.in (L1)	-0.16250	.07730	-2.102	< .0001 ***
	N.out	-0.50733	.07083	-7.163	< .0001 ***
	N.in	-0.32356	.07571	-4.273	< .0001 ***

Monolinguals are

$$\frac{\exp\left(\frac{\Delta E_{mono.E.out} - \Delta E_{mono.G.out} + \Delta G_{mono.G.out} - \Delta G_{mono.E.out}}{2}\right)}{\exp[|\Delta Biling.G.out - \Delta Biling.E.out|]} = \frac{\exp\left(\frac{(0.84580 - 0.06019) - (0.84580 - 0.39887) + (0.69457 - 0.06070) - (0.69457 - 0.18142)}{2}\right)}{\exp[|(0.71302 - 0.16267) - (0.71302 - 0.09162)|]} = 1.17 \quad (8.9)$$

times more susceptible than bilinguals when they see the metaphors out of context. When we test the interaction between speaker group (monolinguals versus bilinguals) and the language specificity of metaphors' idiomatic meanings explicitly using linear mixed effects models, we find that monolinguals' out-of-context ratings are significantly different from bilinguals' ratings, $\chi^2(2) = 32.345$, $p < .0001$.

This difference vanishes when both monolinguals and bilinguals see the metaphor in context. Then, the difference in monolinguals' L1 and L2 ratings decreases to an amount similar to that of bilinguals: for English monolinguals the difference shrinks to

$$|\Delta G.in - \Delta E.in| = |(0.84580 - 0.28695) - (0.84580 - 0.24740)| = 0.03955, \quad (8.10)$$

for German monolinguals to

$$|\Delta G.in - \Delta E.in| = |(0.69457 - 0.09596) - (0.69457 - 0.12443)| = 0.02847; \quad (8.11)$$

similar to bilinguals' rating difference of

$$|\Delta G.in - \Delta E.in| = |(0.71302 - 0.16250) - (0.71302 - 0.13938)| = 0.02312. \quad (8.12)$$

Therefore, monolinguals are

$$\frac{\exp\left(\frac{\Delta E_{mono.E.in} - \Delta E_{mono.G.in} + \Delta G_{mono.G.in} - \Delta G_{mono.E.in}}{2}\right)}{\exp[|\Delta Biling.G.in - \Delta Biling.E.in|]} = \frac{\exp\left(\frac{(0.84580 - 0.24740) - (0.84580 - 0.28695) + (0.69457 - 0.09596) - (0.69457 - 0.12443)}{2}\right)}{\exp[|(0.71302 - 0.16250) - (0.71302 - 0.13938)|]} = 1.01 \quad (8.13)$$

times more susceptible than bilinguals and using the linear mixed effects models to explicitly test for the interaction reveals this difference in odds is not significant, $\chi^2(2) = 3.268$, $p = .1951$.

Notice that context has barely affected bilinguals. Bilinguals' ratings of language-specific metaphors, for which they are familiar with their idiomatic meanings, are unaffected by context: the change in L2 ratings from out-of-context to in-context condition is

$$|\Delta E.in - \Delta E.out| = |(0.71302 - 0.13938) - (0.71302 - 0.09162)| = -0.04776, \quad (8.14)$$

and the change in L1 ratings is even smaller with

$$|\Delta G.in - \Delta G.out| = |(0.71302 - 0.16250) - (0.71302 - 0.16267)| = 0.00017. \quad (8.15)$$

The odds that context would prompt bilinguals' to change their L1 and L2 plausibility judge-

ments are

$$\frac{\exp[|\Delta G.out - \Delta E.out|]}{\exp[|\Delta G.in - \Delta E.in|]} = \frac{\exp[|(0.71302 - 0.16267) - (0.71302 - 0.09162)|]}{\exp[|(0.71302 - 0.16250) - (0.71302 - 0.13938)|]} = 1.05, \quad (8.16)$$

so virtually unaffected by context. However, context has an effect on monolinguals: English monolinguals are

$$\frac{\exp[|\Delta G.out - \Delta E.out|]}{\exp[|\Delta G.in - \Delta E.in|]} = \frac{\exp[|(0.84580 - 0.39887) - (0.84580 - 0.06019)|]}{\exp[|(0.84580 - 0.28695) - (0.84580 - 0.24740)|]} = 1.35 \quad (8.17)$$

times more likely to change their judgement of L2 metaphors relative to L1 metaphors when they see the metaphors in context; German monolinguals are

$$\frac{\exp[|\Delta G.out - \Delta E.out|]}{\exp[|\Delta G.in - \Delta E.in|]} = \frac{\exp[|(0.69457 - 0.06070) - (0.69457 - 0.18142)|]}{\exp[|(0.69457 - 0.09596) - (0.69457 - 0.12443)|]} = 1.10 \quad (8.18)$$

more likely.

Analysis of the interaction between speaker group and the language-specificity of idiomatic meanings revealed that monolinguals' ratings are more susceptible to this language specificity. From here we can move on to the second big question "Do monolinguals and bilinguals make use of context in the same way?" by explicitly testing the interaction between speaker group and context.

The ratings of cross-linguistically shared metaphors change little across experiments and are therefore a good point of comparison. So in testing for the interaction between speaker group and context, the full and reduced linear mixed effects models are constructed on the normalised relative deviation of L2 ratings from those of cross-linguistically shared metaphors for monolinguals. For bilinguals, it is the deviation of both L1 and L2 from shared metaphors, since we already know that their L1 and L2 ratings are similar (difference is 0.07105, out of context, and 0.02312, in context). Analysis revealed a positive interaction between speaker group and context, $\chi^2(1) = 10.83$, $p = .0009986$; monolinguals and bilinguals make different use of context.

Out of context, monolinguals' L2 ratings are 0.216 ± 0.047 (standard errors) lower than those of bilinguals. Context positively interacts with monolinguals, thereby increasing their L2 ratings by 0.185 ± 0.054 . Context barely has an effect on bilinguals; their ratings of language-specific metaphors changed by as little as 0.0024 ± 0.055 . Context helps increase monolinguals' L2 ratings, which makes the L2 ratings more similar to those of L1 metaphors. Bilinguals' L1 and L2 ratings are similar *with* and *without* context.

It would seem that monolinguals are facilitated by the information provided by context during metaphor comprehension. However, since context follows rather than precedes the metaphors, this facilitation most likely comes in the form of re-analysis. Bilinguals seem to hardly be facilitated by the information provided by context at all, potentially because it contributes little new information, i.e. information which, from the perspective of the bilinguals, was not already part of the idiomatic meanings they conventionally associate with the L1 and L2 metaphors. Nevertheless, all of the odds deviate little from 1 and never more than 1.35. We must therefore take the interpretation of these "tendencies" to be preliminary. Further experiments are needed to establish the validity of these "tendencies," which however, is beyond the

Table 8.4: Statistical reliability of the significance of differences in reading/response times. Shown are the number of samples N , the effect size (R^2), the p-value of statistical significance (** for $p \leq .01$, † for $p \leq .1$), the probability of Type I errors α , the probability of Type II errors β , and statistical power $P = 1 - \beta$. The last row in each table provides general estimates of central tendency, averaged across the individual studies and weighted according to their sample size.

Per participant					
	N	Effect size	$p = \alpha$	β	P
E mono	13	.394	.0027 **	.687	.313
Biling	12	.445	.0915 †	.667	.333
	25	.418	.045	.677	.323

Per item					
	N	Effect size	$p = \alpha$	β	P
E mono	25	.394	.0027 **	.250	.750
Biling	30	.445	.0915 †	.083	.917
	55	.422	.051	.159	.841

scope of this thesis and will have to await future research. I suggest that instead we focus on the stronger effect of monolinguals' lowered L2 ratings. Before we conclude this chapter, let us probe the statistical reliability of this effect.

8.6 The reliability of the experimental results

There are two kinds of significance: *substantive* significance and *statistical* significance. Statistical significance is the improbability p (or p-value) that an observed effect could be coincidental. An observed effect could, for instance, be a correlation between an assumed factor of influence (e.g., as in our case, metaphor type, i.e. the language specificity of the non-conceptual aspects of metaphors' idiomatic meanings) and a measurand (e.g. reading/response times). In order to capture assumptions of what coincidental variation in experimental observation would be like, a null model is constructed where the assumed factor of influence is not at play. In our case, we achieved this by deriving a null model from the full linear mixed effects models of variation in reading/response times and plausibility judgements which is "unaware" of a metaphor's type. Following Fisher (1925), the chance of a test statistic to yield false positives, i.e. to make Type I errors, should be no more than 5% or .05 if we are to deem the test reliable. If the p-value falls below this conventionally accepted α threshold, then we may interpret the observed effect to be statistically significant. Improbability p under the null hypothesis is thus an estimate of an α implied by the test statistic, i.e. it is an estimate of the test's true rate of false positives or Type I errors.

Statistical significance is known to be highly sensitive to sample size. Given a sufficiently large number of samples, any effect, no matter how small, may yield a statistically significant p-value, as long as the test's sensitivity (or statistical power) is sufficient, too. Therefore, statistical significance can only be interpreted in light of substantive significance and the test's statistical power. Substantive significance is the size of the observed effect. Cohen (1992)

Table 8.5: Statistical reliability of the significance of differences in out-of-context plausibility judgements. Shown are the number of samples N , the effect size (R^2), the p-value of statistical significance (***) for $p \leq .001$, * for $p \leq .05$, the probability of Type I errors α , the probability of Type II errors β , and statistical power $P = 1 - \beta$. The last row in each table provides general estimates of central tendency, averaged across the individual studies and weighted according to their sample size.

Per participant					
	N	Effect size	$p = \alpha$	β	P
E mono	13	.683	.000057 ***	.201	.799
G mono	19	.475	.022 *	.282	.718
Biling	60	.500	.026 *	.001	.999
	92	.521	.022	.087	.913

Per item					
	N	Effect size	$p = \alpha$	β	P
E mono	25	.683	.000057 ***	.004	.996
G mono	18	.475	.022 *	.320	.680
Biling	30	.500	.026 *	.039	.961
	73	.557	.016	.096	.904

Table 8.6: Statistical reliability of the significance of differences in in-context plausibility judgements. Shown are the number of samples N , the effect size (R^2), the p-value of statistical significance (* for $p \leq .05$, † for $p \leq .1$), the probability of Type I errors α , the probability of Type II errors β , and statistical power $P = 1 - \beta$. The last row in each table provides general estimates of central tendency, averaged across the individual studies and weighted according to their sample size.

Per participant					
	N	Effect size	$p = \alpha$	β	P
E mono	59	.576	.012 *	.001	.999
G mono	20	.443	.063 †	.305	.695
Biling	21	.529	.098 †	.135	.865
	100	.540	.040	.090	.910

Per item					
	N	Effect size	$p = \alpha$	β	P
E mono	26	.576	.012 *	.026	.974
G mono	26	.443	.063 †	.146	.854
Biling	38	.529	.098 †	.004	.996
	90	.518	.063	.051	.949

provides guidelines to evaluate the substantive significance of effect sizes. There are many indices of effect size (e.g. odds ratios, relative risk, Cohen's d , r , or R^2), which, however, can be separated into two groups: (1) measures of absolute difference between groups, also known as the d family (e.g. Cohen's d or Hedges' g), and (2) measures of association, also known as the r family (e.g. the correlation coefficient r or the variation explained through correlation R^2). For the r family of effect size indices, i.e. measures of association and correlation, Cohen suggests that effect sizes of $r \geq .1$ are small, sizes of $r \geq .3$ are medium, and sizes of $r \geq .5$ are large.

Recall that the effect size estimates in the mixed effects models are R^2 . In order to judge them on Cohen's scale we need to take the square root. In the reading/response time experiment where English monolinguals showed slower reading/response times for language-specific German metaphors but not for English or cross-linguistically shared metaphors, the effect of metaphor type at $r = .628$ is large according to Cohen. This effect is statistically significant with $p = .0027$ (see Table 8.4). If the results had not achieved significance, this would already be indicative of the absence of a real effect. Statistical power tells us that the effect is only reliably visible in the per-item analysis ($P > .8$), which means that more participants would be needed to meet statistical quality standards for a reliable per-participant analysis. The per-item analysis, which is already statistically reliable, nevertheless already suggests that we are dealing with a real effect. Recall that I said in Section (7.4) in Chapter 7 that the effect of metaphor type on reading/response times was originally not apparent in an analysis of variance. It was only when I redid the analysis using linear mixed effects models that the effect became clear. The result of the original ANOVA result, at the time, gave reason not to continue trials with German monolingual native speakers. However, from the power analysis we see now that further testing would be advisable.

In the two experiments, where English and German monolingual native speakers judged metaphors' plausibility outside of context and found L2 language-specific metaphors less plausible than L1 and cross-linguistically shared metaphors, the effect of metaphor type with $r_E = .826$ and $r_G = .689$ is also large. The effect of metaphor type observed for English and German monolinguals is also statistically significant with $p_E = .000057$ and $p_G = .022$ (see Table 8.5). Weighted averages for prior and posterior probability estimates indicate that the effect of metaphor type is more than 90% reliable. In the experiments where speakers rated metaphors' plausibility in context and monolingual speakers showed lower ratings for L1 metaphors, similar to bilinguals, the effect of metaphor type is large for all three speaker groups ($r_E = .759$, $r_G = .666$, $r_B = .727$, see Table 8.6). However, the effect is only below $\alpha = .05$ for English monolinguals ($p_E = .012$); for German monolinguals and bilinguals, it is below $.1$ ($p_G = .063$ and $p_B = .098$). Nevertheless, statistical power is $.854$, well above the value of $.8$ recommended by Cohen (1988). The reliability metrics thus suggest that, given a larger number of participants, statistical significance *would* fall below $.05$.

8.7 Summary

In the previous chapter, we saw that while monolinguals judge L1 language-specific and cross-linguistically shared metaphors to be plausible, they perceive L2 language-specific metaphors to be significantly less plausible. Bilinguals, on the other hand, also gave lower plausibility ratings to L1-specific metaphors and I argued that this is how bilinguals' sensitivity for metaphors' language specificity is reflected in their use of the rating scale in the judgement task. As part

of a context creation task, participants were asked to come up with sensible continuations to two L1, two L2, two shared, and two novel metaphors. The most common continuations that monolinguals came up with were selected as contexts for the experiments in this chapter. Continuations for the other 42 metaphors in the experiments were added, also with the intention to improve plausibility. CMT, RT, and GS agree that sensible, helpful context, which contributes new information to the discourse, should improve plausibility ratings. Thus the central question of the experiments in this chapter was: If context improves the intelligibility of metaphors across languages, do bilinguals make use of context in the same way that monolinguals do?

CMT, RT, and GS agreed that monolinguals' in-context ratings of plausibility for L2 and novel metaphors should increase compared to the out-of-context condition, which was confirmed by the experimental findings. This finding suggests that their conceptual knowledge, which we independently know to be cross-linguistically maximally similar, is not enough to make L2 language-specific metaphors cross-linguistically as communicative as cross-linguistically shared metaphors. This contradicts CMT's assumption that metaphor is primarily a cognitive-conceptual phenomenon and not a linguistic one. CMT, on the one hand, and RT and GS, on the other, disagreed when it came to monolinguals' ratings of the L1 and cross-linguistically shared metaphors. Recall that CMT takes linguistic intuitions such as plausibility judgements to be rather direct evidence of the mental representation of meaning salience (*conceptual metaphors*), while RT and GS doubt that intuitions directly reflect conceptual representation and instead claim that linguistic intuitions reflect speakers' implicit or explicit knowledge of non-conceptual linguistic conventions (e.g. relevance expectations and Fregean senses). Given monolinguals' high plausibility ratings of L1 and shared metaphors, CMT predicted that these ratings should be unaffected because context should not contribute something beyond what is already part of speakers' *conceptual* mental representation of the idiomatic meanings of these metaphors. RT and GS, on the other hand, predicted that context would affect L1 and shared metaphors. Experimental findings confirmed RT and GS's view and falsified CMT's view.

CMT, RT, and GS agreed that bilinguals' plausibility ratings of novel metaphors should also increase with context, because for novel metaphors, bilinguals are in the same position as monolinguals: there are no conventionally established idiomatic meanings for the novel metaphors which monolinguals or bilinguals could employ as inferential shortcuts and so both monolinguals and bilinguals must consider all plausible inferential pathways during comprehension, led by their conceptual knowledge, Fregean senses, intuitions of meaning salience, and expectations of contextual relevance. However, CMT, on the one hand, and RT and GS, on the other, disagreed on how context would affect bilinguals' ratings of L1 metaphors. Since CMT takes plausibility judgements to rather directly reflect speakers' conceptual knowledge and since bilinguals' plausibility ratings of L1 metaphors were significantly lower than cross-linguistically shared metaphors in the out-of-context condition, CMT predicted that the plausibility of L1 metaphors should increase with context. RT and GS predicted that, given speakers' high expectations of contextual relevance for metaphors with highly conventional salient idiomatic meanings, any context that would *not* meet these high expectations should be suboptimal and result in decreased plausibility. Experimental findings showed that bilinguals' ratings of both L1 and L2 language-specific metaphors were lower than those of cross-linguistically shared metaphors. Through post-experimental interviews we know that this second group of bilingual participants was as familiar with the idiomatic meanings of the language-specific metaphors as the group of bilinguals in the out-of-context condition. So bilinguals' plausibility judge-

ments, similar to monolinguals' judgements, do not directly reflect conceptual knowledge, but knowledge of non-conceptual linguistic conventions.

We defined metaphors' cross-linguistic communicability in two ways: L2 language-specific metaphors are strongly communicative if they are as communicative as L1-specific and cross-linguistically shared metaphors and they are weakly communicative if they are at least significantly more communicative than unfamiliar novel metaphors. However, we saw two problems with these definitions in the in-context experiments: (1) for monolinguals, the plausibility ratings of L1 metaphors decrease with context to the point where they are more similar to those of L2 metaphors. We therefore must rephrase strong communicability as: L2 language-specific metaphors are strongly communicative if they are as communicative as cross-linguistically shared metaphors. (2) Since the plausibility judgements of both of the control conditions, shared and novel metaphors, are affected by context, they are not "ideal" control conditions. I argued that we should look at the rate of change of language-specific metaphors as a function of the rate of change of the shared and novel metaphors, the change being the change in plausibility ratings from the out-of-context to the in-context condition. It turned out that monolinguals' plausibility judgements are more affected by context than bilinguals' judgements, and this effect is greater for metaphors with language-specific idiomatic meanings than it is for non-language-specific metaphors. Therefore, monolinguals and bilinguals make use of context differently. Monolinguals use the contexts of L2 language-specific metaphors to inform inferential comprehension of these metaphors very much in the same way that they use context to inform their interpretation of unfamiliar novel metaphors for which they have no idiomatic meanings. In doing so they must solely rely on their conceptual knowledge, Fregean senses, expectations of salience and contextual relevance in a distinctly non-idiomatic, compositional way. It must be that the contexts of language-specific metaphors do not provide bilinguals with information they did not already have and their knowledge of the idiomatic meanings results in bilinguals having stricter expectations of contextual relevance for language-specific metaphors than monolinguals.

Chapter 9

Summary and conclusions

9.1 Introduction

In the previous two chapters, we saw that monolinguals' comprehension of L2 metaphors compared to L1 and cross-linguistically shared metaphors is affected by context, whereas bilinguals' rating behaviour is not. Out of context, monolinguals showed lower reading/response/reaction times and lower plausibility ratings for L2 than L1 or cross-linguistically shared metaphors. In context, monolinguals did not show slower reading/response/reaction times and lower plausibility ratings for L2 metaphors. Bilinguals' rating behaviour is unaffected by context. Context must therefore provide monolinguals, but not bilinguals, with information which they facilitate during L2 metaphor comprehension. I suggested that bilinguals may ignore L2 metaphors' context because the information it provides is *not new* to them, whereas it is *new* to monolinguals. The contexts were intended to improve L2 metaphor plausibility, which they did. Subject to the deficits in the material of achieving conceptual cross-linguistic similarity, the information that contexts provided and monolinguals benefited from could be both a function of conceptual and a function non-conceptual aspects of metaphors' language-specific idiomatic meanings. In order to gauge which of the two aspects are the more likely cause, we need to view the experimental estimates of their effect size in light of the conceptual deficits in the experimental material. In Chapter 6, we determined these conceptual deficits pre-experimentally through cross-linguistic Levenshtein distance of language-specific Fregean senses, while controlling for contextual relevance and frequency (Chapter 5). In this chapter, we summarise the experimental findings for monolingual and bilingual speakers respectively and conclude that conceptual aspects of metaphors' idiomatic meanings play a more important role than non-conceptual aspects during comprehension.

9.2 Theoretical background

A metaphor, such as *time is money*, consists of three things: (1) a 'source' or 'vehicle' concept, MONEY, (2) a 'target' or 'tenor' concept, TIME, and (3) an 'implied analogy' that invites us to comprehend the target concept in terms of the source concept. Historically, there are generally three views on what concepts are: (1) they are mental representations of sensory perception, experience out of interaction with the world, and imagination. (2) Concepts are best understood through our ability to use them. If the successful completion of a specific task

requires that we have a particular concept, then a demonstration of our ability to complete the task should be indicative of us having that concept. (3) The view that concepts are Fregean sense relations, i.e. that they are abstract symbols whose meaning is defined in the propositional system of a language and by conditions of truth in the world. The physiological and behavioural evidence supports the mental representations view most and its modern version has become known as Embodied Cognition. On this view all basic primary concepts in the conceptual system (speakers' mental lexicons) are rooted in perceptual experience or introspective sensory experience of body-internal states such as emotions.

To comprehend a metaphor is to infer under which conditions the analogy implied between the source concept MONEY and the target concept TIME holds. This often means that not all properties of the source concept are transferred to the target concept, but only a subset of them (*semantic narrowing*), e.g. one Fregean sense of the concept TIME is that of musical metre, which, however, is not relevant in the context of the analogy implied between MONEY and TIME. In order to facilitate the analogy, properties may also be inferred that cannot be said to be part of either the source or the target concept, called *emergent properties* (Gineste et al., 2000): e.g. PET FISH has the property 'brightly coloured' which cannot be said to be a property of either PET or FISH. The complex meaning of the metaphor that satisfies the conditions of the implied analogy is thus *pragmatically enriched* and more than the mental contents of its constituent source and target concepts. A common way of reconciling pragmatic enrichment with the idea that concepts have core meanings is to think of the inferential process as a process of deriving an *ad hoc* concept from the source concept such that its properties match the target features of the target concept, and it is this *ad hoc* concept and not the original concept that is attributed to or included in the class, category, or domain represented by the target concept (McGlone, 1996; Glucksberg, McGlone & Manfredi, 1997; Glucksberg, Manfredi & McGlone, 1997). The pragmatically enriched set of inferences necessary to satisfy the metaphoric analogy, together with the relevant properties of the source and target concepts (i.e. their *ad hoc* concepts), are a linguistic meaning, specifically *speaker meaning*, the context-dependent meaning implied by the speaker and the one that is most likely to be inferred by other interlocutors.

According to Relevance Theory (RT) (Sperber & Wilson, 1986; Wilson & Sperber, 1993; Sperber & Wilson, 1995; Wilson & Sperber, 2002, 2004) interlocutors draw those inferences during inferential comprehension that are cognitively most-relevant to understanding the meaning which the speaker is most likely to imply in a given discourse, as well as understanding the speaker's communicative intent in doing so. Assumptions about which information is cognitively relevant to interlocutors is constantly updated as the discourse progresses. Linguistic meaning is thus highly dependent on the context in which it is uttered, the information gathered from the preceding discourse, interlocutors' conceptual and mental lexicons, and the speaker's communicative intent that the interlocutors infer. RT rejects the Gricean view that comprehending metaphors requires dedicated inferential machinery (Grice, 1975, 1978) and instead claims that there is one set of inferential mechanisms that form the machinery to comprehend any and all linguistic utterances. On this view, inferring a metaphoric meaning differs from other linguistic speaker meanings in the kinds of inferences that have to be drawn in order to satisfy the implied analogy. Pilkington (2000) claims that the inferences drawn in metaphor comprehension are more likely to be a large set of weak implicatures rather than few strong inferences (e.g. entailments or strong implicatures) and these weak implicatures accumulate to produce the rich

mental imagery associated with the metaphoric analogy in question.

According to Graded Saliency (GS) (Giora, 1997, 1998, 1999, 2003; Peleg et al., 2008; Peleg & Giora, 2011) the inferential process consists of two modules: (1) a linguistic module that infers salient meanings based on linguistic knowledge and (2) a pragmatic module that enriches the meaning by taking non-linguistic knowledge such as conceptual, experiential, perceptual, contextual, and world knowledge into consideration. The more familiar, common, conventional, and prototypical a meaning is, the more likely it is to come to mind more easily and readily, and the more salient and likely to be inferred the meaning is. Inferential strength, on this view, is then a function of *meaning saliency*. According to GS, the linguistic meaning that is ultimately inferred for an utterance is one that is a compromise between the outputs of the two modules. The two modules are thought to run in parallel so that processing may align at any point during comprehension. This alignment is necessary in order to optimally arrive at a compromise. On this view, metaphor comprehension requires no dedicated cognitive mechanisms beyond the machinery of the two modules and, similar to RT, that an inferred meaning happens to be metaphoric is then a result of the inferences drawn and not the mechanisms involved. Since the second module allows context to be considered, the linguistic meaning that is inferred by the two modules working in unison is *speaker meaning*. GS's view of metaphoric meanings is thus compatible with that of RT.

Conceptual Metaphor Theory (CMT) (Lakoff & Johnson, 1980; Lakoff, 1983, 1987; Lakoff & Johnson, 1999, 2003; Lakoff, 2008, 2009) claims complex inferences, such as linguistic meanings, are the result of conceptual mappings from primary embodied concepts to ever more complex higher-order concepts. Which mappings are established and activated to yield complex inferences, according to CMT, depends on sensory-perceptual stimuli (context, in RT and GS terms) and the mental content and structure of concepts based on past experiences. Although CMT thinks of linguistic meanings as the result of interconnecting primary concepts to ever more complex higher-order concepts, this view is compatible with RT and GS's view of *speaker meaning* as the context-dependent linguistic meaning implied by the speaker and inferred by listeners. The complex inferences that complex higher-order concepts yield for a particular context, in CMT terms, are the complex inferences that correspond to context-dependent speaker meanings, in RT and GS terms. CMT takes metaphoric-analogical reasoning to be a structural feature of the mental architecture of the conceptual system. Comprehending metaphors, on this view, thus comes natural to the human mind.

Interpretations of metaphoric analogies can become conventionalised as, for instance, in the case of metaphoric proverbs such as *time is money*. A conventional linguistic meaning is also called *idiomatic meaning* because similar to idioms, the full, pragmatically enriched meaning of the metaphoric analogy is not inferable from the linguistic form of the proverb and speakers' conceptual knowledge of the source and target concept alone. Speakers are provided with the missing information through linguistic convention and the missing information can be thought of as the context that would have to be 'spelled out' in order to allow interlocutors to infer the full, pragmatically enriched meaning in a non-idiomatic way. The idiomatic meaning of a metaphoric proverb can thus be understood as an inferential shortcut: instead of having to consider all conceptually possible alternative interpretations and assessing their plausibility in the context of the analogy implied by the metaphor, speakers who are familiar with the idiomatic meaning are provided with a default interpretation, which eliminates the need to consider all alternatives. The idiomatic meaning therefore is a salient and preferred meaning.

Researchers in cognitive science (Lakoff & Johnson, 1980; Lakoff, 1983, 1987; Coulson, 2006) have collected substantial evidence of cross-linguistic systematicity in linguistic metaphors, which they take as evidence for (a) the metaphoric structure of the conceptual systems of the respective languages and (b) the mental representation of the respective linguistic metaphors. They claim that the structure of the conceptual system is inherently metaphoric-analogical and that therefore conceptual interconnections are also metaphor-like. The evidence of systematic patterns in linguistic metaphors, on this view, is taken to be indicative of the kinds of metaphoric conceptual interconnections that exist in the respective languages and these interconnections are therefore called *conceptual metaphors*. On this view, conceptual metaphors are preferred, salient meanings. They may be universal to us as a species or specific to a particular language, language family, or culture. However, many language-specific linguistic metaphors have conventional meanings called *idiomatic meanings* associated with them. On CMT's view, metaphoric idiomatic meanings are then language-specific conceptual metaphors. Importantly, this characterises idiomatic meanings as language-specific ways of conceptualising perceptual experience and introspective mental states. Contrary to CMT, RT and GS claim that idiomatic meanings are not primarily motivated by conception but by the cooperative need (cf. Grice's Cooperative Principle) to facilitate communication. On CMT's view, metaphoric idiomatic meanings constitute conceptual mental representations. On RT and GS's view, metaphoric idiomatic meanings require knowledge of linguistic conventions that are distinctly *non-conceptual* because they are not primarily motivated by conceptual plausibility, world knowledge, or conceptual knowledge. With respect to metaphoric idiomatic meanings, the difference between the opposing claims made by CMT, on the one hand, and RT and GS, on the other, thus is how much influence metaphoric idiomatic meanings have on the ways that speakers conceptualise the world around them: CMT takes the influence to be rather great and largely unconscious; RT and GS take the influence to be indirectly mitigated by non-conceptual linguistic conventions. We can rephrase this with respect to the evidence of linguistic metaphors: CMT takes the systematicity in linguistic metaphors to be rather direct evidence of the cognitive processes involved and of the mental structures that these processes create; RT and GS take the insight the linguistic evidence gives into cognition to be rather indirect and mitigated by linguistic conventions.

This thesis seeks to experimentally gauge the amount of influence of non-conceptual linguistic conventions in metaphoric idiomatic meanings. In order to do this we need to set up a situation where speakers of two languages cross-linguistically have similar conceptual knowledge with regards to the source and target concepts of particular linguistic metaphors, similar expectations of the contextual relevance of their conceptual properties and Fregean senses, and similar mental imagery. We then need to identify linguistic metaphors that have idiomatic meanings which are conventionalised in only one but not the other language. Metaphoric proverbs have language-specific conventional meanings which will only be fully intelligible to non-native speakers if they are able to infer a meaning similar to the one that native speakers have through convention (the idiomatic meaning). If CMT is correct, then non-native speakers should be able to infer meanings for the metaphoric proverbs that are similar to the idiomatic meanings that native speakers have. If RT and GS are correct, then the meanings that non-native speakers infer for the proverbs should be different from the idiomatic meanings of native speakers and the magnitude of the difference should correspond to the size of the effect that non-conceptual linguistic conventions have on idiomatic meanings.

9.3 The experimental rationale

If inferential comprehension is primarily a conceptual process, as CMT assumes, then when speakers have similar concepts pertaining to metaphors' language-specific idiomatic meanings, they should also comprehend these metaphors similarly. On CMT's view, a metaphoric proverb's idiomatic meaning is a complex concept inferred from its source and target concept by employing a conceptual metaphor that has been conventionalised specific to a particular language. When two languages cross-linguistically share the embodied concepts and conceptual metaphors necessary to infer the meaning of a particular metaphoric proverb, then non-native speakers should be able to infer the same complex concept as native speakers. In doing so, it should suffice for speakers to be led by their conceptual knowledge alone to draw plausible inferences about metaphors with L2 language-specific idiomatic meanings.

If the idiomatic meaning of a metaphoric expression, in RT and GS terms, is a linguistic meaning that is inferred on the basis of the source and target concept and based on the given context or a context that is associated with it out of considerations of inferential-cognitive relevance, but, critically, this linguistic meaning is not inferable without knowledge of non-conceptual linguistic conventions, conventions that are not motivated by or subject to conceptual plausibility, then speakers who are unaware and unfamiliar with these non-conceptual conventions should be unable to infer meanings for metaphoric expressions that are approximately similar to their idiomatic meanings. In other words, if idiomatic meanings require knowledge of linguistic conventions that have nothing to do with how speakers conceptualise but which only constrain how they communicate these ideas (concepts), then in a cross-linguistic situation where we know that speakers of two language communities conceptualise in similar ways (with respect to the concepts of particular metaphoric expressions) speakers who do not know the relevant L2 language-specific linguistic conventions should infer metaphoric meanings that are distinctly and significantly different from the true idiomatic meanings that speakers familiar with the conventions infer. If idiomatic meanings require knowledge of particular linguistic conventions, then when speakers do not have this knowledge, they should be unable to infer these idiomatic meanings, i.e. they should infer other plausible linguistic meanings.

The experiments in this thesis set up a situation where the chances of cross-linguistic intelligibility of metaphoric proverbs should be maximised conceptually so that if, as CMT proposes, cross-linguistic metaphor comprehension is primarily dependent on cross-linguistic approximate conceptual similarity, then language-specific metaphoric proverbs should be cross-linguistically as intelligible as cross-linguistically shared proverbs in this situation. If, however, cross-linguistic metaphor comprehension is not primarily dependent on metaphoric conception but approximate similarity of non-conceptual linguistic conventions concerning proverbs' idiomatic meanings, as RT and GS propose, then cross-linguistic intelligibility should be lower for language-specific metaphoric proverbs than for cross-linguistically shared ones even when cross-linguistic approximate conceptual similarity is maximised for all metaphors in question.

In investigating the language specificity of idiomatic meanings of metaphoric proverbs, there are two control conditions when looking at two languages in close linguistic contact: (1) metaphoric proverbs for which speakers of both languages know idiomatic meanings and these meanings are the same for the same proverbs (*cross-linguistically shared* metaphoric proverbs), and (2) *novel* metaphoric expressions whose linguistic form resembles that of the proverbs but for which speakers of neither language have any idiomatic meanings because

no such meanings have been conventionally established yet. The two control conditions are theory-neutral: Whatever the nature of idiomatic meanings is, be it language-specific conceptual metaphors or language-specific linguistic non-conceptual conventions, all proverbs categorized as ‘cross-linguistically shared’ in this thesis are classified as such because they share idiomatic meanings (in addition to shared linguistic form, commonness of and familiarity with that form, and lexical and compositional semantics). Metaphoric expressions classified as ‘novel’ have not yet been conventionally established and thus neither speakers of English nor German can have any idiomatic meanings for them.

The dependent variable, the cross-linguistic intelligibility of idiomatic metaphoric expressions, is measured in three ways in this thesis: (1) through reading/response times, (2) through metaphor plausibility judgements, and (3) through a context creation task. The independent variables are: (1) participants’ language proficiency (English monolinguals, German monolinguals, English-German bilinguals), (2) the language specificity of metaphoric idiomatic meanings (English-specific, German-specific, cross-linguistically shared, and unconventionalised-novel), and (3) the mode of presentation (out of context and in context).

The rationale of measurand 1, reading/response times: If knowledge of the idiomatic meaning acts as an inferential shortcut, native speakers, who have this knowledge, but not non-native speakers, because they lack this knowledge, should be faster at reading and judging metaphoric proverbs which require knowledge of idiomatic meanings. CMT would predict that because of the closeness of language contact between English and German and the exchange of concepts and conceptual metaphors that comes with it, when native speakers of either English or German make use of the same basic concepts and conceptual metaphors for a particular proverb, that proverb should be cross-linguistically intelligible to non-native speakers; in other words, non-native speakers should be able to infer a meaning similar to the one that native speakers have through convention (the idiomatic meaning). Non-native speakers should therefore not show slower reading/response times for proverbs they do not know from their native language, but read them as fast as proverbs that are cross-linguistically shared between English and German. RT and GS, on the other hand, would predict that if the idiomatic meaning is not purely a matter of conception but requires knowledge of linguistic convention, then non-native speakers who lack this knowledge should be unable to use the inferential shortcut that native speakers have through their knowledge of the idiomatic meaning and non-native speakers should therefore take longer comprehending L2 language-specific proverbs than L1 and shared ones because they have to consider all conceptually plausible interpretations while native speakers simply default to the idiomatic meaning.

The rationale of measurand 2, metaphor plausibility judgements: Metaphors are communicative if they are intelligible and in order to be intelligible they must first be plausible. Therefore, a metaphor whose implied analogy between source and target is implausible is also not intelligible and not communicative. The analogy is plausible only if speakers feel justified making the necessary assumptions as part of pragmatic enrichment. Metaphor plausibility judgements can therefore be viewed as an index of their intelligibility and communicativeness. Because of the closeness of language contact between English and German and the concepts and conceptual metaphors that are shared between them because of this closeness, when speakers of both languages share the basic concepts and conceptual metaphors for a particular metaphoric proverb, that proverb should be as intelligible to non-native speakers as it is to native speakers and this, CMT would predict, should be true by virtue of conceptual plausibility. RT and GS,

however, would predict that the proverb should not be cross-linguistically intelligible on the basis of conception alone, but only if speakers know the linguistic conventions associated with it: the idiomatic meaning, omitted context, and expectations of contextual relevance. RT and GS would thus predict that non-native speakers should not find proverbs from the other language as plausible as proverbs from their native language because their considerations of ‘plausibility’ include their knowledge of linguistic conventions (or lack thereof) *on top of* purely conceptual considerations and they would predict that considerations of linguistic convention should take precedence over purely conceptual considerations.

Meaning is well-known to be highly context-dependent and this is recognised by all three theories, CMT, RT and GS. In this thesis, metaphor plausibility judgements are therefore measured in and out of context. Context affects the plausibility judgements of the two control conditions. Since measurand 2 defines weak and strong intelligibility relative to the two control conditions, we need to evaluate weak and strong intelligibility as the rate of change in plausibility ratings of language-specific metaphors *as a function of* the rate of change in the plausibility ratings of the control conditions, the change being the change from out-of-context to in-context metaphor presentation. CMT, RT, and GS agree when speakers should facilitate contextual information during comprehension: if it yields positive cognitive effects, is informative, contributes something new to the discourse, aids comprehension, is an explanation of conceptual metaphors (i.e. it is a substitute for the conceptual metaphors that speakers might lack), and more context should make it easier to form *ad hoc* concepts. In other words, context is relevant to comprehension if it is helpful during comprehension. If contextual information is relevant for comprehending a metaphor, then speakers should make use of it. If contextual information is irrelevant for comprehending a metaphor or conceptually conflicts with what speakers already inferred, speakers should ignore this new contextual information. With enough relevant contextual information, cross-linguistically unintelligible metaphors should become intelligible.

The rationale of measurand 3, contextual continuations: If comprehending a particular metaphor requires knowledge of its idiomatic meaning and the idiomatic meaning can be thought of as the context omitted from overt linguistic expression that would be necessary to ‘spell out’ the full conventional speaker meaning in a non-idiomatic way, then native speakers who know the intended full conventional meaning should be able to make (at least part of) the omitted context explicit. Since the idiomatic meaning also includes expectations as to which contextual information would be helpful during comprehension, when speakers are asked to create sensible context for a metaphoric proverb, only native speakers but not non-native speakers should be able to meet these conventional expectations. If it is appropriate to think of idiomatic meanings as language-specific conceptual metaphors, then non-native speakers should be able to make helpful context explicit on the basis of their conceptual knowledge and considerations of conceptual plausibility whenever they share the necessary basic concepts and conceptual metaphors with native speakers. CMT would propose that idiomatic meanings are language-specific conceptual metaphors and would therefore predict that non-native speakers should be able to do this for two languages where, for the proverbs in question, basic concepts and conceptual metaphors are cross-linguistically shared because of intense language contact. RT and GS, on the other hand, see the idiomatic meaning, the knowledge of omitted context, and expectations of helpful context as part of arbitrary linguistic convention and not conception, and not necessarily as motivated by Embodied Cognition. RT and GS would therefore predict that non-native speakers should fail to make context explicit which is relevant to the conventional

meaning of an L2 proverb.

9.4 The degree to which maximising cross-linguistic approximate conceptual similarity has been achieved

We want to test whether non-native speakers are able to infer the same complex metaphor concepts that native speakers have through convention (the *idiomatic* meaning). Hence, we want to test if non-native speakers can infer these meanings for metaphors that are unfamiliar to them. Monolingual native speakers of English should be presented with German-specific metaphoric proverbs that are unfamiliar to them because the proverbs do not occur in English. Conversely, monolingual German native speakers should be presented with English-specific proverbs that are unfamiliar to them because the proverbs do not occur in German. Bilingual speakers of English and German, however, should be familiar with all English-specific and German-specific proverbs, and thus serve as a control group. Cross-linguistically shared proverbs that occur in both English and German should be familiar to all monolingual and bilingual speakers. Metaphoric expressions that resemble the form of the proverbs but which do not occur at all, neither in English nor German, should be unfamiliar to all monolingual and bilingual speakers. However, English and German have one of the closest language contacts in the world leading to many concepts being cross-linguistically shared. It is therefore much harder to find proverbs that occur in only one of the two languages but not the other. Through an extensive analysis of corpora of written and transcribed spoken English and German, totalling ~300 billion words worth of text for English and ~40 billion words for German, I identified metaphoric proverbs that are specific to one but not the other language and two controls: metaphoric proverbs that occur in both languages and metaphoric novel expressions that do not occur at all, but resemble the other proverbs.

I showed how to convert the raw corpus frequency of a particular proverb into the *probability* of speakers with a certain knowledge of vocabulary (i.e. one indicator of language proficiency) to be familiar with that proverb. This operationalisation of *familiarity* uses the notion of *salience* as defined by Giora (1997, 1998, 1999, 2003); Peleg et al. (2008); Peleg & Giora (2011). The probability of the English-specific and German-specific proverbs identified through the corpus analysis varied by as little as 6%. The probability of cross-linguistically shared proverbs varied even less between English and German, at about 4.5%. We also want to make sure that the language-specific and cross-linguistically shared proverbs have been in use long enough so that experimental participants of both languages have had an equal chance to learn and familiarise themselves with them. The average participant was 33 years old. I therefore suggested that we look at the corpus frequency of the proverbs 30 years ago and check whether there have been any drastic changes compared to their present frequency. Out of the 50 proverbs selected for the experiments, only two cross-linguistically shared proverbs and one language-specific proverb had seen a significant change in frequency. The two shared proverbs *time is money* and *knowledge is power* had increased in frequency. The only language-specific proverb that had changed significantly in frequency was the English proverb *a home is a castle*. It had increased in frequency as well. None of the metaphoric proverbs had seen any drastic decreases. Diachronic frequency changes for all language-specific and cross-linguistically shared proverbs differed only 4% between English and German. I demonstrated how we can ensure that the synchronic and

diachronic frequency counts are reliable and accurate by measuring the frequency variation of punctuation marks, which are less likely to change over time than other parts of the grammar. For the past 200 years this variation was remarkably small and although there was a change in the use of punctuation between 1850 and 1900, before and after this change the average frequency remained stable. Frequency variation of punctuation marks thus seems to be a good estimate of the measurement accuracy of a corpus. This also allowed us to check whether those proverbs categorised as ‘language-specific’ really are language-specific: a reason why we found a particular proverb only in corpora of one language but not the other (therefore being classified as a language-specific proverb) might be that the corpora of the other language were too small to find infrequent occurrences. I showed that the frequency of the language-specific proverbs was larger than the variation of measurement accuracy. The occurrence of proverbs in corpora of one but not the other language (and their categorisation as language-specific as a result) is therefore unlikely to be an artefact of corpus size. The corpus frequency of the words in the metaphoric proverbs and novel expressions were even more frequent than the most frequent proverb and differed even less between corpora of the two languages than the frequency of proverbs as a whole. They differed by as little as 1.3%.

Cross-linguistic differences in the form of metaphoric proverbs and novel expressions was 9.6%, on average, but most of this was caused by differences in morphology, pronunciation, and the presence of articles in noun phrases: German has a morphology that is richer than that of English. In the metaphors, the number of morphemes differed by 5.6% between the two languages. Most of the morphological mismatches were cases such as *silence/Schweigen*. In the grammar of English, *silence* is a noun that consists of only one morpheme, the stem, here, whereas *Schweigen*, in the grammar of German, is a noun derived from a verb and the verb consists of a stem *schweig-* and an infinitive ending *-en*. Hence, *Schweigen* has two morphemes and *silence* only one. Many of the words in the proverbs are historically related to the same stem forms (cognates), e.g. *life/Leben*, *seeing/Sehen*, *love/Liebe*, *blind/blind*, *revenge/Rache* (*wrath* is cognate to *Rache*), *volcano/Vulkan*, but a lot of times letters that are silent in English, e.g. the *e* in *love*, are pronounced in German, e.g. the final *e* in *Liebe*. The sounds that differed were mostly vowels, which means that this also led to there being more syllables in German than there are in the English version of the material. For instance, *love* has one syllable in English, but *Liebe* has two syllables in German. Pronunciation differences amounted to around 20%, but they would have no influence on speakers’ interpretation of the metaphors since they are solely a result of systematic differences between the phonological systems of the two languages. The presence of an article in a noun phrase, however, might have an influence on speakers’ inferences. Consider, for instance, the difference in the meaning of *love* between *the love we share* and *love is all around*. But these differences cancelled each other out between metaphor-proverb types and between the two languages, and even so, amounted to a cross-linguistic difference of only 10%, some of which were also cases such as systematic grammatical differences in the mass-count noun distinction between the two languages, e.g. *bliss* does not allow an article in the proverb *ignorance is bliss* while *Segen* has to have an article in the German equivalent of the cross-linguistically shared proverb *Unwissenheit ist ein Segen*. Therefore, most of the cross-linguistic difference of 9.6% in the form of the metaphoric proverbs and novel expressions was due to systematic grammatical differences with little effect on metaphors’ interpretation and not due to differences in speakers’ familiarity with the language-specific and cross-linguistically shared metaphors.

I showed that when we look at the number of different senses of the words used in the proverbs and novel expressions and compare that number between English and German as an index of the difference in lexical meanings, we find that the cross-linguistic difference amounted to only 16% and these 16% of differing senses were only 10% of the time likely to have an effect on the interpretation of metaphoric proverbs (i.e. 1.6% overall). We can consider the copula (a form of the verb *be*, either *is* or *are*, in English, and *ist* or *sind*, in German) to be semantically vacuous. We can thus say that each metaphor consists of two lexical meanings, the lexical meanings of the metaphor source and target concept. When we then add up the number of sense differences of the two lexical meanings per metaphoric proverb, we saw that there was a cross-linguistic difference of again around 16%; hence no more than when we consider lexical items individually. Apart from speakers' knowledge of language-specific idiomatic meanings, the semantics of the metaphors thus differed by only 14% between English and German. The semantic differences, thus, were cross-linguistically as large as the systematic grammatical differences and larger than the differences in familiarity. However, remember that for the cross-linguistically shared proverbs there are conventionally accepted forms which we have to use in order for speakers to recognise the idiomatic form. And for the language-specific proverbs, the words I chose in their translation was the best choice compared to lexical alternatives. For all intents and purposes the proverbs and novel expressions were thus cross-linguistically comparable in terms of their linguistic form, the familiarity of that form, and their lexical semantics. Hence, the only significant difference between them should be that monolinguals will only know idiomatic meanings for proverbs they know from their native language while bilinguals should know idiomatic meanings for all proverbs that have been classified as language-specific according to the corpus analysis. When cross-linguistic differences in the material were unavoidable, these differences were evenly distributed across metaphor types and can thus not explain the systematic differences in plausibility judgements we see in the experiments.

The degree to which maximising cross-linguistic approximate conceptual similarity has been achieved can be quantified as follows: Taking book translations as evidence for the magnitude of the exchange of concepts as cultural ideas, English and German have a closer language contact with each other than with any other languages (see Figure 4.2, Chapter 4, Section 4.4.4). Closeness of language contact is thus maximised. The salience and conventionality of proverbs' idiomatic meanings and cross-linguistic approximate similarity of conceptual alternatives is maximised based on the corpus analysis and evidenced in the context creation task by the finding that monolingual native speakers and bilinguals more frequently allude to the idiomatic meanings in their continuations than to conceptual alternatives, but when they allude to conceptual alternatives, these alternatives are cross-linguistically similar for both native and non-native speakers. The corpus analysis also shows that the familiarity of metaphors' linguistic form is maximised. Metaphors' source and target concepts, conceptual alternative compositional meanings and the expected contextual relevance of those alternative meanings are cross-linguistically approximately similar. The linguistic form of the language-specific metaphoric proverbs in the experiments is 82.15% cross-linguistically similar, which is comparable to the cross-linguistic similarity of 85.3% for cross-linguistically shared and novel metaphors. The linguistic meaning of the language-specific metaphoric proverbs in the experimental material is 83.75% cross-linguistically similar, which is comparable to the cross-linguistic semantic similarity of 88.55% for shared and novel metaphors. Cross-linguistic similarity of the linguistic form and meaning of language-specific metaphors is thus maximised and about as high as for

cross-linguistically shared and novel metaphors.

9.5 Experimental findings

CMT claims that idiomatic meanings of language-specific metaphoric proverbs are language-specific conceptual metaphors that together derive the complex concept which corresponds to the idiomatic meaning from basic primary embodied concepts. RT and GS claim that language-specific idiomatic meanings are not necessarily only language-specific ways of conceptualising, but can also include non-conceptual knowledge provided through linguistic convention. If CMT's claim regarding the nature of metaphoric idiomatic meanings is correct, then for two languages such as English and German and metaphors such as the language-specific metaphoric proverbs, where the metaphors' source and target concepts, expectations of contextual relevance, and associations with other mental imagery are cross-linguistically approximately similar, monolingual native speakers of English, for instance, should be able to infer linguistic meanings for language-specific German proverbs that are similar to the idiomatic meanings that German native speakers know. Similarly, monolingual German native speakers should be able to infer meanings for English-specific metaphors that are cross-linguistically approximately similar to the idiomatic meanings that English native speakers infer. Bilinguals serve as a control group because, contrary to monolingual native speakers of either English or German, they should be familiar with the idiomatic meanings of both English and German language-specific proverbs as well as all cross-linguistically shared proverbs, and they should therefore be able to use inferential shortcuts for all language-specific metaphors in the same way as for cross-linguistically shared metaphors. Finding that the cross-linguistic communicability of L2 language-specific metaphors is lowered for monolingual English and German native speakers would thus be evidence in favour of RT and GS and evidence against CMT, and the magnitude of decreased communicability, if it was found, should correspond to the size of the effect that non-conceptual linguistic conventions have on inferential comprehension of language-specific metaphoric proverbs.

Cross-linguistic communicability of L2 language-specific metaphors is lowered for monolinguals

Cross-linguistic intelligibility and communicability of L2 language-specific metaphoric proverbs is lowered for monolinguals despite maximised cross-linguistic approximate conceptual similarity of metaphors' source and target concepts, contextual relevance expectations, cross-linguistic similarity of speakers' familiarity with the conventional form and meaning of metaphors, and closeness of language contact. Given in Table (9.1) are the p-values for lowered L2 metaphor intelligibility when metaphors are presented out of context. The effect size R^2 given in the table is the variance explained by the fixed effect *metaphor type* in the linear mixed effects models. Recall that metaphor type, compared to the random effects in the mixed models, is the best predictor of reading/response times and plausibility judgements. Monolinguals read and responded more slowly for L2-specific metaphors than for L1-specific or shared metaphors (the p-values given in Table 9.1 are the pairwise comparisons between L2 and shared metaphors, the lower p-value, and between L1 and L2 metaphors, the higher p-value; these p-values were calculated using linear mixed effects models and using RTs, not RTs converted to z-scores).

Monolinguals and bilinguals took between 2165.5ms and 4213ms to read, judge the plausibility of metaphors, and respond. This makes them conscious but still intuitive judgements. Monolingual native speakers read and respond significantly slower to L2 metaphors than to L1 metaphors ($p = .0027$, see Table 7.3) or cross-linguistically shared metaphors ($p = .0000082$, see Table 7.3), but bilinguals did not (RT difference between German and English metaphors: $p = .5308$, German and shared metaphors: $p = .2849$). Monolinguals took about 1 second longer to read and judge metaphors unfamiliar to them (L2 language-specific and novel ones) than familiar ones (L1 and cross-linguistically shared metaphors). Bilinguals took around half a second longer to judge language-specific metaphors than non-language-specific ones.

Table 9.1: Summary of the first experimental finding. Cross-linguistic intelligibility of L2 language-specific metaphors is lowered for monolinguals. Shown are the number of sample N, the effect size (R^2), the p-value of statistical significance, the probability of type I errors α , weighted average probability of type II errors β , and average statistical power $\bar{P} = 1 - \beta$, weighted by the sample size of the original study.

Measurand	N	Effect size	$p = \alpha$	β	\bar{P}
Reading/response times	38	.394	.0000082 - .0027	.399	.601
Plausibility judgements	75	.475 - .683	.000057 - .022 ^a	.119	.881
Context creation	32		significant ^b	NA	NA

Given in Table (9.1) are the out-of-context plausibility judgements (table footnote a) because the reading/response times are also for the out-of-context condition. Cross-linguistic metaphor intelligibility, as evidenced by plausibility judgements, increases for L2 language-specific metaphors when monolingual speakers see them in context, but monolinguals still perceive them to be less plausible than L1-specific and cross-linguistically shared metaphors ($.012 \leq p \leq .063$). When a second group of monolingual English native speakers saw the metaphoric proverbs and novel metaphors *with* the contextual continuations that the first group of German native speakers had created, they rated language-specific German proverbs to be *more plausible* than the first group of English speakers who saw the proverbs without contextual continuations. The presence of context had no influence on how monolingual German native speakers judged the plausibility of English-specific proverbs. Recall that the plausibility judgements of cross-linguistically shared proverbs and novel metaphors serve as control conditions relative to which we evaluate the plausibility judgements of language-specific proverbs. Because the presence of context affects the two control conditions, I argued that we need to evaluate the *rates of change* in plausibility judgements and not just the *absolute* ratings.

We then saw that context has systematic effects on language-specific, cross-linguistically shared, and novel metaphors: (1) when speakers are aware of the relevance of idiomatic meanings for particular metaphors, they have very high expectations as to what constitutes a plausible context, and (2) when they do not know of idiomatic meanings for particular metaphors, their contextual expectations are lowered and any context that seems conceptually plausible is likely to be facilitated in comprehending these metaphors. (3) Monolinguals' plausibility judgements are more affected by context than bilinguals' judgements, and (4) this effect is two to three times greater for metaphors with language-specific idiomatic meanings than it is for non-language-specific metaphors. Bilinguals' plausibility judgements are less affected by the presence or absence of context than monolinguals. Monolinguals do not know the idiomatic

meanings of L2 language-specific proverbs and thus fully rely on context to inform their judgements of conceptual plausibility. Bilinguals know idiomatic meanings for the language-specific proverbs and thus do not rely on context and thus their judgements of plausibility of language-specific metaphors do not change from out-of-context to in-context presentation. But it also shows that while monolinguals understand ‘plausibility’ in the task to be *conceptual* plausibility for L2-specific proverbs, bilinguals understand it as *linguistic* plausibility, i.e. for them considerations of contextual relevance, linguistic knowledge of the idiomatic meaning, and familiarity of the idiomatic form *outweigh* purely conceptual considerations. This shows that whenever linguistic knowledge is available and relevant during comprehension, it takes precedence over purely conceptual intuitions.

In the context creation task, participants were asked to create sensible contextual continuations to 8 of the 50 metaphors, so as to improve their plausibility. Analysis of the metaphoric senses that participants alluded to in these continuations revealed that plausibility judgements in the first task were not lower for L2 proverbs because of conceptual differences. On the contrary, in the continuations we saw a lot of conceptual agreement between speakers of the two languages; only the idiomatic meanings are not cross-linguistically intelligible because non-native monolingual speakers did not allude to them in any of their continuations. Monolinguals do not allude to meanings for L2 metaphors in the context creation task that are similar to the idiomatic meanings that L2 native speakers and bilinguals know and allude to (table footnote b). The meanings that monolinguals allude to for L2 metaphors are similar to the alternative meanings that L2 native speakers and bilinguals allude to whenever they do not allude to the idiomatic meaning. This shows that native and non-native speakers are drawing inferences based on the same conceptual knowledge and that the cross-linguistic approximate similarity of that knowledge is indeed maximised with respect to the metaphors in question. However, this conceptual knowledge does not seem to allow monolinguals to infer meanings for L2 metaphors that are similar to their idiomatic meanings. When we compare the continuations created by English and German monolinguals we see the following: they create similar continuations for cross-linguistically shared metaphors that imply the same sets of senses and they are primarily conventional idiomatic senses and secondarily conceptually plausible senses. When they create continuations for language-specific metaphors, only the continuations of native speakers allude to the conventional idiomatic sense of the metaphor while non-native speakers create continuations that are conceptually sensible but which diverge from the idiomatic meaning. For novel metaphors speakers of both languages create continuations that are conceptually plausible and we find that they come to the same conceptually plausible senses.

We can conclude two things: (1) whenever conventional idiomatic meanings are known to speakers (as in the case of cross-linguistically shared and L1 language-specific metaphors), monolingual speakers prefer them over meanings that are just conceptually plausible, and (2) when monolingual speakers do not associate any conventional idiomatic meaning with a particular linguistic metaphor (as in the case of novel and L2 language-specific metaphors), they rely primarily on conceptually plausible interpretations. We can therefore conclude that whenever conventional linguistic meanings are available to speakers, they take precedence over, i.e. they are more salient than pure considerations of conceptual plausibility. L2 language-specific metaphors are cross-linguistically intelligible between the two languages in close contact only in the sense that non-native speakers come to the same conceptual interpretations as native speakers, but non-native speakers are *unable* to infer those meanings that native speakers consider

conventional idiomatic meanings.

The same line of reasoning holds true for the control group of bilinguals: (1) they know conventional idiomatic meanings for cross-linguistically shared as well as all language-specific metaphors and so these are the meanings that their continuations allude to most often, more often than continuations that allude to merely conceptually plausible inferences. (2) For novel metaphors, bilinguals are in the same position as monolinguals: they do not know of any conventional meanings, hence their inferences focus on conceptually salient features of metaphor source and target concepts and on drawing further analogies from the implied metaphoric analogy. Therefore, bilinguals come up with continuations for novel metaphors that are very similar to those created by monolinguals and we see that the range of senses they allude to is limited to the same set of concepts and conceptual mappings that monolinguals used to construct continuations for the novel metaphors. This suggests that, apart from bilinguals' knowledge of idiomatic meanings for the L2 metaphors, bilinguals seem to make use of the same inventory of concepts and conceptual mappings as the English and German monolinguals when they make inferences about the metaphors. It would thus seem that monolinguals' lower plausibility ratings for L2 metaphors could be solely attributed to them being unfamiliar with the necessary idiomatic meanings. Bilinguals, on the other hand, who also give lower ratings to L2 metaphors seem to do so as an expression of their awareness of metaphors' language-specificity because their continuations show an awareness of the idiomatic meanings and that awareness was also evident in post-experimental interviews.

On all three measurands, the differences between L2 language-specific metaphors, on the one hand, and L1-specific and cross-linguistically shared metaphors, on the other, are significantly greater than we should expect according to CMT. The findings thus falsify CMT's predictions on all three measurands and support the predictions made by RT and GS. It is therefore more likely that idiomatic meanings are language-specific non-conceptual linguistic constraints on conception than that they are language-specific conceptual metaphors. Native speakers, who know the idiomatic meaning, default to an appropriate interpretation of the metaphor while non-native speakers, who are unfamiliar and unaware of the idiomatic meaning, are left to rely on their conceptual and linguistic knowledge of their own language. If the lexical meaning, sense relations, and connotations of a metaphor source or target concept then are sufficiently different from the other language in their native language, then these differences in linguistic knowledge are enough to mislead non-native speakers away from the interpretation that native speakers reach by following the idiomatic meaning. As we saw through the plausibility judgements and the contextual continuations in the experiments, the idiomatic meaning is not in and of itself the conceptually most plausible interpretation; native speakers only think so because they default to it out of linguistic convention. It is therefore understandable that non-native speakers, unaware of the necessary idiomatic meanings, are left to rely purely on conceptual considerations and, being led by reason and, importantly, the linguistic conventions from their native language, they reach a different metaphoric interpretation. The conventional metaphoric meaning that native speakers reach for the language-specific proverbs is often not the conceptually most plausible one, which is why non-native speakers do not feel justified to make the necessary assumptions during pragmatic enrichment even if they might consider them. The inferences that non-native speakers would have to draw to make the necessary assumptions, to them, are much weaker implicatures than they are for native speakers.

Bilinguals' sensitivity to metaphors' language specificity

Bilinguals are sensitive to the target language of language-specific metaphoric proverbs and consequently treat them differently from cross-linguistically shared proverbs in experimental tasks. We might take this to suggest that the mental representation of their idiomatic meanings could also be different from that of the idiomatic meanings of shared proverbs. While context affected monolinguals' plausibility judgements, bilinguals' judgements were not affected. It is likely that bilinguals were not affected because context did not provide them with any information that they felt necessary to incorporate during inferential processing so as to infer a more plausible interpretation of the metaphoric analogies. In the context creation task we see that bilinguals do indeed know the idiomatic meanings of all language-specific and cross-linguistically shared proverbs because their continuations allude to the idiomatic meanings most commonly inferred (Table 9.2, footnote b). Post-experimental interviews also revealed that bilinguals were well aware during experiments that some of the metaphoric proverbs were language-specific and others were non-specific, and reading/response times showed much less delay for language-specific metaphors than the delays we saw for monolinguals and L2-specific metaphors. We would therefore have expected bilinguals to judge language-specific metaphoric proverbs to be as plausible as cross-linguistically shared ones. Nevertheless, they give lower plausibility judgements to language-specific metaphors. I think it is therefore reasonable to conclude that bilinguals use the rating scale in the plausibility judgement task quite differently from monolinguals. For monolinguals, plausibility judgements correlate with familiarity: metaphors that are familiar to them (cross-linguistically shared and L1 metaphors) receive high plausibility ratings, those that are unfamiliar (L2 and novel metaphors) receive low plausibility ratings. The presence of context has a twofold effect on monolinguals' judgements: (1) unfamiliar L2 language-specific metaphors receive higher plausibility ratings than novel metaphors, and (2) the plausibility of familiar metaphors decreases in the presence of context because of the very strict expectations of contextual relevance and this decrease is more pronounced for L1 than shared metaphors, most likely because the relevance expectations are stricter.

For bilinguals, on the other hand, plausibility judgements reflect their awareness and sensitivity to metaphors' language specificity: they give the highest plausibility ratings to cross-linguistically shared metaphors and even though the presence of context decreases the plausibility of shared metaphors, this decrease is as large as the decrease we see for monolinguals' plausibility ratings of shared metaphors. If it is indeed the metaphor-specific expectations of contextual relevance that cause the decrease, then monolinguals and bilinguals seem to have similar relevance expectations for shared metaphors. Bilinguals give the lowest plausibility ratings to novel metaphors because, like monolinguals, they have and associate no idiomatic meanings with them that could serve as default interpretations. Bilinguals are therefore in the same position as monolinguals and must consider all alternative interpretations during inferential processing to choose the most plausible one, and, like monolinguals, the most plausible meanings that they infer for novel metaphors are still less plausible than for metaphors for which they know idiomatic meanings. The presence or absence of context has no effect on how bilinguals perceive the plausibility of language-specific metaphors, but notice that bilinguals' plausibility ratings for language-specific metaphors are as high as monolinguals' ratings for L1 metaphors in the in-context condition. I think it is therefore reasonable to assume that context has no effect on bilinguals' ratings of language-specific metaphors because the conflicts with context

that lead monolinguals to give decreased plausibility ratings to L1 metaphors must already be incorporated in bilinguals' relevance expectations for the language-specific metaphors.

We can think of bilinguals' sensitivity to metaphors' language specificity as an extension of expectations of contextual relevance because knowing which languages are potential target languages for using a particular metaphor is, in a way, contextual knowledge: if we consider the two languages that the English-German bilinguals speak to constitute a set of potential discourses with a certain potential to use particular metaphors, then shared metaphors, because they occur in both languages, can occur in more discourse contexts than metaphors that are restricted to only one of the two languages. In that sense, bilinguals' expectations of the contextual relevance of language-specific metaphors are more restrictive than those of monolinguals. Monolinguals, on the one hand, are unfamiliar with the L2 metaphors and consequently cannot know they are different from the novel metaphors and, on the other hand, they are unaware that the L1 metaphors are restricted to their native language and consequently do not know they are different from the shared metaphors. Therefore, bilinguals' medial plausibility ratings express their sensitivity to metaphors' language-specificity, while for monolinguals the twofold effect of context on L1 and L2 metaphors only makes it seem as if monolinguals and bilinguals were considering the same knowledge during inferential processing of language-specific metaphors. So even though in-context presentation makes for more naturalistic experiments that simulate natural language discourse more realistically, had we only looked at monolinguals' and bilinguals' plausibility judgements in context, then we might have been under the false impression that monolinguals' and bilinguals' inferential metaphor comprehension worked quite similarly when in fact it does not.

Table 9.2: Summary of the second experimental finding. Bilinguals treat language-specific metaphors (L) differently from cross-linguistically shared metaphors (S) in the plausibility judgement task. Shown are the number of sample N, the effect size (R^2), the p-value of statistical significance, the probability of type I errors α , weighted average probability of type II errors $\bar{\beta}$, and average statistical power $\bar{P} = 1 - \beta$, weighted by the sample size of the original study.

Measurand	N	Effect size	$p = \alpha$	$\bar{\beta}$	\bar{P}
Reading/response times	42	.445	.092 – .285	.250	.750
Plausibility judgements	90	.4997 – .529	.026 – .303 ^a	.028	.972
Context creation	60		significant ^b	NA	NA

Table (9.2) summarises bilinguals' performance on the the three measurands. The effect size R^2 given in the table is the variance explained by the fixed effect *metaphor type* in the linear mixed effects models. Recall that metaphor type, compared to the random effects in the mixed models, is the best predictor of reading/response times and plausibility judgements. Notice that the p-values for reading/response times are not significant. It would thus seem that reading/response times are not able to pick up on bilinguals' sensitivity to metaphors' language specificity or this sensitivity is not reflected in reading/response times. Plausibility judgements, however, seem to reveal bilinguals' sensitivity to metaphors' language specificity. Note that in Table (9.2), the higher p-value is insignificant (table footnote a). However, this is a rather simplified view. When metaphors were presented out of context, German proverbs received significantly lower ratings than shared metaphors (G~S $p = .026$) but English proverbs did not (E~S $p = .303$). When bilinguals saw the metaphors in context, they rated both German

(G~S $p = .079$) and English proverbs (E~S $p = .098$) to be significantly less plausible than shared ones. All bilinguals saw the English language-specific proverbs in German translation, but those tested in Edinburgh judged their plausibility more similar to shared proverbs (E~S $p = .303$) than those tested in Germany (E~S $p = .098$). The most likely explanation is probably a sociolinguistic bias depending on whether they were tested in an English-speaking country, where English proverbs would be more salient and frequent in everyday use, than in a German-speaking country, where bilinguals would almost never be exposed to them and the proverbs' idiomatic meanings would thus be much less salient when bilinguals are tested. Those bilinguals tested in Edinburgh might have been socio-linguistically primed to expect English metaphors and consequently were more likely to give them higher plausibility ratings than bilinguals tested in Germany.

9.6 General conclusion

According to the three ways in which the cross-linguistic communicability of L2 language-specific metaphoric proverbs was tested in the experiments in this thesis, non-native speakers infer meanings for these proverbs that are distinctly different from the idiomatic meanings that native speakers default to out of linguistic convention, and this is true even when cross-linguistic approximate conceptual similarity of metaphors' source and target concepts, similarity of contextual relevance expectations, and closeness of language contact are maximised for the proverbs in question. It would thus seem that inferring their idiomatic meanings requires language-specific knowledge of linguistic conventions that are distinctly non-conceptual because they are not motivated by conceptual plausibility, world knowledge, or conceptual knowledge. Taking the metaphoric image schemas in these language-specific proverbs to be indicative of conception would be to mis-classify non-conceptual linguistic knowledge as conceptual knowledge. If we did, we would falsely assume a correlation between the linguistic form of these proverbs and the complex concepts that are their idiomatic meanings to be primarily an insight into speakers' way of conceptualising; rather it is an insight into non-conceptual linguistic convention.

Bibliography

- Barr, D. J., Levy, R., Scheepers, C., & Tilly, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.
- Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211–227.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.0-6. URL: <http://CRAN.R-project.org/package=lme4>.
- Bayart, A. (1958). La correction de la logique modale du premier et second ordre s5. *Logique et Analyse*, *1*, 28–44. Translated in Cresswell forthcoming.
- Bayart, A. (1959). Quasi-adéquation de la logique modale de second ordre s5 et adéquation de la logique modale de premier ordre s5. *Logique et Analyse*, *2*, 99–121.
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, *53*, 370–418.
- Bennett, M. & Hacker, P. (2008). *History of Cognitive Neuroscience*. Oxford: Wiley-Blackwell.
- Berger, J. O. & Jeffreys, W. H. (1991a). Minimal Bayesian testing of precise hypotheses, model selection, and Ockham’s razor. Technical Report, Purdue University.
- Berger, J. O. & Jeffreys, W. H. (1991b). Sharpening Ockham’s razor on a Bayesian stop. Technical Report, Purdue University.
- Blasko, D. G. & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 295–308.
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127–135.
- Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H. J., & Jancke, L. (2005). Scanning silence: mental imagery of complex sounds. *Neuroimage*, *26*, 1119–1127.
- Cao, P., Maximov, A., & Südhof, T. C. (2011). Activity-dependent IGF-1 exocytosis is controlled by the Ca²⁺-sensor Synaptotagmin-10. *Cell*, *145*(2), 300–311.

- Carnap, R. (1937). *The Logical Syntax of Language*. New York: Humanities.
- Carnap, R. (1942). *Introduction to Semantics*. Cambridge, Mass.: Harvard University Press.
- Carnap, R. (1947). *Meaning and Necessity*. Chicago: The University of Chicago Press.
- Carruthers, P. (2000). *Phenomenal Consciousness: A Naturalistic Theory*. New York: Cambridge University Press.
- Carruthers, P. (2006). *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*. New York: Oxford University Press.
- Carston, R. (2002). Metaphor, ad hoc concepts and word meaning – more questions than answers. *UCL Working Papers in Linguistics*, 14.
- Chiappe, D., Kennedy, J. M., & Smykowski, T. (2007). Reversibility, aptness, and the conventionality of metaphors and similes. *Metaphor and Symbol*, 18(2), 85–105.
- Chomsky, N. (1996). *Powers and Prospects*. Boston MA: South End Press.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- Colston, H. L. & Gibbs, R. W. (2002). Are irony and metaphor understood differently? *Metaphor and Symbol*, 17, 57–80.
- Coulson, S. (2006). Metaphor and Conceptual Blending. In *Encyclopedia of Language & Linguistics* (2nd ed.). (pp. 32–39).
- Creem-Regehr, S. H. & Lee, J. N. (2005). Neural representations of graspable objects: are tools special? *Cognitive Brain Research*, 22, 457–469.
- Damásio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam Publishing.
- Davies, M. (2004). BYU-BNC (Based on the British National Corpus from Oxford University Press). Available online at <http://corpus.byu.edu/bnc/>.
- Davies, M. (2008). The Corpus of Contemporary American English: 450 million words, 1990–present. Available online at <http://corpus.byu.edu/coca/>.
- Davies, M. (2010). The Corpus of Historical American English: 400 million words, 1810–2009. Available online at <http://corpus.byu.edu/coha/>.
- Davies, M. (2011a). Google Books (American English) Corpus (155 billion words, 1810–2009). Available online at <http://googlebooks.byu.edu/>.
- Davies, M. (2011b). Google Books (British English) Corpus (155 billion words, 1810–2009). Available online at <http://googlebooks.byu.edu/>.

- Davies, M. (2011c). Google Books Corpus. (Based on Google Books n-grams). Available online at <http://googlebooks.byu.edu/>. Based on: Jean-Baptiste Michel*, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Koseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden*. Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* 331 (2011).
- de Saussure, F. (1916). *Cours de linguistique générale*. In C. Bally & A. Sechehaye (Eds.), *Linguistics: The Cambridge Survey. Vol. 1. Linguistic Theory: Foundations*. Glasgow: Fontana/Collins. With the collaboration of A. Riedlinger, Lausanne and Paris: Payot; trans. W. Baskin, *Course in General Linguistics*, 1977.
- de Saussure, F. (1993). Saussure's Third Course of Lectures in General Linguistics (1910–1911). In E. Komatsu & R. Harris (Eds.), *Emile Constantin ders notlarindan, Language and Communication Series, Volume 12*. Oxford: Pergamon. Trans. and ed. E. Komatsu and R. Harris.
- Decety, J. (1996). Do imagined and executed actions share the same neural substrate? *Cognitive Brain Research*, 3, 87–93.
- Descartes, R. (1641). *Meditationes de prima philosophia*. Amsterdam: Apud Ludovicum Elzevirium.
- Dudenredaktion (2013). *Duden - Die deutsche Rechtschreibung* (26nd ed.). Mannheim: Dudenverlag.
- Dummett, M. (1993). *Seas of Language*. Oxford: Oxford University Press.
- Farah, M. J. (2004). *Visual Agnosia* (2nd ed.). Cambridge, MA: MIT Press.
- Faul, F., Buchner, A., Erdfelder, E., & Lang, A.-G. (2014). G*Power. Available online at <http://www.gpower.hhu.de/>. Version used: 3.1.9.2.
- Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd. Chapter 5, p 102.
- Fodor, J. (1975). *The Language of Thought*. Cambridge MA: Harvard University Press.
- Fodor, J. (1983). *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge MA: MIT Press.
- Fodor, J. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge MA: MIT Press.
- Fodor, J. (1990). *A Theory of Content and Other Essays*. Cambridge MA: MIT Press.
- Fodor, J. (1991). Replies. In G. Rey & B. Loewer (Eds.), *Meaning in Mind: Fodor and His Critics*. Blackwell, Oxford.
- Fodor, J. (1994). *The Elm and the Expert: Mentalese and Its Semantics*. Cambridge MA: MIT Press.

- Fodor, J. (2008). *LOT 2: The Language of Thought Revisited*. New York: Oxford University Press.
- Fodor, J. & Lepore, E. (1996). The red herring and the pet fish: why concepts still can't be prototypes. *Cognition*, *58*, 253–270.
- Frege, G. (1879). *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle a. S.: Louis Nebert.
- Frege, G. (1892a). On sense and reference. In P. Geach & M. Black (Eds.), *Translations from the Philosophical Writings of Gottlob Frege* (pp. 56–78). Blackwell, Oxford. Translation published 1952.
- Frege, G. (1892b). Über Begriff und Gegenstand. *Vierteljahresschrift für wissenschaftliche Philosophie*, *16*, 192–205. Translated as Concept and Object, by P. Geach in *Translations from the Philosophical Writings of Gottlob Frege*, P. Geach and M. Black (eds. and trans.), Oxford: Blackwell, third edition, 1980.
- Frege, G. (1892c). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, *100*, 25–50.
- Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Cognitive Brain Research*, *20*(2), 226–241.
- Gauss, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis Solem ambientium*. (*Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections*). New York: Dover. English translation by C. H. Davis.
- Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., & Fried, I. (2008). Internally generated reactivation of single neurons in human hippocampus during free recall. *Science*, *322*(5898), 96–101.
- Gerardin, E., Sirigu, A., Lehéricy, S., Poline, J.-B., Gaymard, B., Marsault, C., Agid, Y., & Le Bihan, D. (2000). Partially overlapping neural networks for real and imagined hand movements. *Cereb. Cortex*, *10*(11), 1093–1104.
- Gibbs, R. (1992). Categorization and metaphor understanding. *Psychological Review*, *99*, 572–577.
- Gibbs, R. (1994a). Figurative Thought and Figurative Language. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 411–446). Academic Press, San Diego.
- Gibbs, R. (1994b). *The Poetics of Mind: Figurative Thought, Language and Understanding*. Cambridge: Cambridge University Press.
- Gibbs, R. (1996). Why many concepts are metaphorical. *Cognition*, *61*, 309–319.
- Gibbs, R. (1998). The fight over metaphor in thought and language. In N. Katz, C. Cacciari, R. Gibbs, & M. Turner (Eds.), *Figurative Language and Thought* (pp. 88–118). Oxford University Press, New York.
- Gibbs, R. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics*, *34*, 457–486.

- Gibbs, R. (2003). Embodied experience and linguistic meaning. *Brain and Language*, 84, 1–15.
- Gibbs, R. (2006). *Embodiment and cognitive science*. New York: Cambridge University Press.
- Gibbs, R. & Steen, G. (Eds.). (1999). *Metaphor in cognitive linguistics*. Amsterdam: John Benjamins.
- Gineste, M., Indurkha, B., & Scart, V. (2000). Emergence of features in metaphor comprehension. *Metaphor and Symbol*, 15, 117–135.
- Giora, R. (1997). Understanding figurative and literal language: the graded salience hypothesis. *Cognitive Linguistics*, 7, 183–205.
- Giora, R. (1998). Discourse coherence is an independent notion: A reply to Deirdre Wilson. *Journal of Pragmatics*, 29, 75–86.
- Giora, R. (1999). On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics*, 31, 919–929.
- Giora, R. (2003). *On our Mind: Salience, Context and Figurative Language*. New York: Oxford University Press.
- Giora, R., Gazal, O., & Goldstein, I. (2012). Salience and Context: Interpretation of Metaphorical and Literal Language by Young Adults Diagnosed with Aspergers Syndrome. *Metaphor and Symbol*, 27, 22–54.
- Glucksberg, S. & Keysar, B. (1993). How Metaphors Work. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 401–424). Cambridge University Press, Cambridge.
- Glucksberg, S., Manfredi, D., & McGlone, M. (1997). Metaphor comprehension: How metaphors create new categories. In T. Ward, S. Smith, & J. Vaid (Eds.), *Creative Thought: An Investigation of Conceptual Structures and Processes* (pp. 327–350). American Psychological Association, Washington DC.
- Glucksberg, S., McGlone, M. S., & Keysar, B. (1992). Metaphor Understanding and Accessing Conceptual Schema: Reply to Gibbs (1992). *Psychological Review*, 99(3), 578–581.
- Glucksberg, S., McGlone, M. S., & Manfredi, D. (1997). Property Attribution in Metaphor Comprehension. *Journal of Memory and Language*, 36, 50–67.
- Google (2012). Google Books Ngram Viewer datasets. Available online at <http://storage.googleapis.com/books/ngrams/books/datasetsv2.html>.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech arts* (pp. 41–58). Academic Press.
- Grice, H. P. (1978). Further notes on logic and conversation. In P. Cole (Ed.), *Syntax and semantics 9: Pragmatics* (pp. 113–128). Academic Press.
- Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge MA: Harvard University Press.
- Groenewold, R., Bastiaanse, R., Nickels, L., Wieling, M., & Huiskes, M. (2015). The differential effects of direct and indirect speech on discourse comprehension in Dutch and English listeners with and without aphasia. *Aphasiology*, 29, 685–704.

- Gunther, Y. (2003). *Essays on Nonconceptual Content*. Cambridge, MA: MIT Press.
- Hampton, D. W., Innes, N., Merkler, D., Zhao, C., Franklin, R. J. M., & Chandran, S. (2012). Focal immune-mediated white matter demyelination reveals an age-associated increase in axonal vulnerability and decreased remyelination efficiency. *American Journal of Pathology*, *180*(5), 1897–1905.
- Hebb, D. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley.
- Heck, R. G. (2000). Nonconceptual content and the space of reasons. *Philosophical Review*, *109*, 483–523.
- Hintikka, J. (1957). Modality as referential multiplicity. *Ajatus*, *20*, 49–64.
- Hintikka, J. (1961). Modality and Quantification. *Theoria*, *27*, 119–128.
- Hume, D. (1739). *A Treatise of Human Nature*.
- Institut für Deutsche Sprache (2013). COSMAS II corpora and search interface. Available online at <http://www.ids-mannheim.de/cosmas2/>.
- Jaynes, E. T. (1974). *Probability Theory with Applications in Science and Engineering. A Series of Informal Lectures*. Posthumous manuscript, placed online <http://bayes.wustl.edu/etj/science.pdf.html>.
- Jeffreys, H. (1939). *Theory of Probability* (1st ed.). Oxford: The Clarendon Press.
- Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 20–49). New York: Routledge.
- Jones, G. V. (1982). Stacks not fuzzy sets: An ordinal basis for prototype theory of concepts. *Cognition*, *12*, 281–290.
- Jones, L. L. & Estes, Z. (2006). Roosters, robins, and alarm clocks: Aptness and conventionality in metaphor comprehension. *Journal of Memory and Language*, *55*(1), 18–32.
- Kahneman, D. (2002). Maps of bounded rationality: A perspective on intuitive judgment and choice. In *Nobel Prize in Economics, Prize Lecture*, (pp. 459–464).
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.
- Kay, P. & Kempton, W. (1984). What is the Sapir-Whorf-hypothesis? *American Anthropologist*, *86*, 65–79.
- Kelly, S. D. (2001). The non-conceptual content of perceptual experience: Situation dependence and fineness of grain. *Philosophy and Phenomenological Research*, *62*(3), 601–608.
- Kenny, A. (2010). Concepts, Brains, and Behaviour. *Grazer Philosophische Studien*, *81*(1), 105–113.
- Keysar, B., Shen, Y., Glucksberg, S., & Horton, W. S. (2000). Conventional language: How metaphorical is it? *Journal of Memory and Language*, *43*, 576–593.

- Kövecses, Z. (1990). *Emotion concepts*. New York: Springer Verlag.
- Kövecses, Z. (2000). *Metaphor and emotion*. Cambridge and New York: Cambridge University Press.
- Kövecses, Z. (2002). *Metaphor: a practical introduction*. Oxford: Oxford University Press.
- Kripke, S. (1959). A Completeness Theorem in Modal Logic. *Journal of Symbolic Logic*, 24(1), 1–14.
- Kripke, S. (1963a). Semantical Analysis of Modal Logic I: Normal Modal Propositional Calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 9, 67–96.
- Kripke, S. (1963b). Semantical Considerations on Modal Logic. *Acta Philosophica Fennica*, 16, 83–94.
- Kripke, S. (1972). *Naming and Necessity*. Cambridge, MA: Harvard, 1980.
- Lakoff, G. (1983). Cognitive semantics: Some recent developments. *Mathematical Social Sciences*, 4(2), 183–184.
- Lakoff, G. (1987). *Women, fire, and dangerous things: what categories reveal about the mind*. Chicago: University of Chicago Press.
- Lakoff, G. (2008). The neural theory of metaphor. In R. Gibbs (Ed.), *Cambridge Handbook of Metaphor and Thought* (pp. 17–38). Cambridge University Press, New York.
- Lakoff, G. (2009). The neural theory of metaphor. <http://dx.doi.org/10.2139/ssrn.1437794>.
- Lakoff, G. (2013). Embodied Cognition and Language. Talk given at the Central European University on Neural Theory, October 2013. Available online at <https://www.youtube.com/watch?v=XWYaoAoijdQ>.
- Lakoff, G. (2014). How Brains Think. Talk given at the 4th International Conference on Foreign Language Teaching and Applied Linguistics, May 2014, Sarajevo. Available online at <https://www.youtube.com/watch?v=1dDAfoVdYU8>.
- Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: the embodied mind and its challenge to Western thought*. New York: Basic Books.
- Lakoff, G. & Johnson, M. (2003). *Metaphors we live by (new edition with 'Afterword')*. Chicago: University of Chicago Press.
- Lamport, L. (1994). *L^AT_EX: A document preparation system* (2nd ed.). Reading MA: Addison-Wesley Publishing Company. Illustrations by Duane Bibby.
- Langacker, R. W. (1982). Space Grammar, Analysability, and the English Passive. *Language*, 58(1), 22–80.
- Langacker, R. W. (1987a). *Foundations of Cognitive Grammar, Volume 1, Theoretical Prerequisites*. Stanford: Stanford University Press.

- Langacker, R. W. (1987b). *Foundations of Cognitive Grammar, Volume 2, Descriptive Application*. Stanford: Stanford University Press.
- Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. New York: Oxford University Press.
- Laplace, P.-S. (1812). *Théorie Analytique des Probabilités*. Paris: Courcier.
- Leibniz, G. W. (1686). Discours de métaphysique (Discourse on Metaphysics).
- Leibniz, G. W. (1704). *Nouveaux Essais sur L'entendement humain (New Essays on Human Understanding)*. Cambridge: Cambridge University Press. Translated by Peter Remnant and Jonathan Bennett in 1981.
- Leibniz, G. W. (1714). Principes de la nature et de la grâce fondés en raison (Principles of Nature and of Grace). In C. I. Gerhardt (Ed.), *Die philosophischen Schriften* (7 vols ed.). Hildesheim: Georg Olms. Reprinted 1965.
- Lerman, H. (2010). Non-conceptual experiential content and reason-giving. *Philosophy and Phenomenological Research*, 81(1), 1–23.
- Levelt, W. J. M. (1989). *Intention to Articulation*. Cambridge, MA: MIT Press.
- Lewis, D. (1969). *Convention: A Philosophical Study*. Cambridge MA: Harvard University Press.
- Lewis, D. (1975). Languages and Language. In K. Gunderson (Ed.), *Minnesota Studies in the Philosophy of Science*, volume VII. University of Minnesota Press, Minnesota.
- Lewis, D. (1978). Truth in fiction. *American Philosophical Quarterly*, 15(1), 37–46.
- Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, 8, 339–359.
- Lewis, D. (1983). Languages and Language. In D. Lewis (Ed.), *Philosophical Papers*, volume I (pp. 163–188). Oxford University Press, Oxford.
- Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Basil Blackwell.
- Li, P., Abarbanell, L., Gleitman, L., & Papafragou, A. (2011). Spatial reasoning in tenejapan mayans. *Cognition*, 120, 33–53.
- Locke, J. (1690). *An Essay Concerning Human Understanding*.
- Locker, L., Hoffman, L., & Bovaird, J. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavioral Research Methods*, 39, 723–730.
- Margolis, E. (1994). A reassessment of the shift from the classical theory of concepts to prototype theory. *Cognition*, 51, 73–89.
- Margolis, E. & Laurence, S. (1999). *Concepts: Core Readings*. Cambridge MA: MIT Press.
- Margolis, E. & Laurence, S. (2010). Concepts and theoretical unification. *Behavioral and Brain Sciences*, 33, 219–220.

- Margolis, E. & Laurence, S. (2011). Concepts. The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), Edward N. Zalta (ed.), URL: <http://plato.stanford.edu/archives/spr2014/entries/concepts/>.
- McDermott, M. (1988). The narrow semantics of names. *Mind*, 97(386), 224–237.
- McGlone, M. S. (1996). Conceptual metaphors and figurative language interpretation: Food for thought? *Journal of Memory and Language*, 35, 544–565.
- McGlone, M. S. (2007). What is the Explanatory Value of a Conceptual Metaphor? *Language and Communication*, 27, 109–126.
- McGonigle, D. J., Howseman, A. M., Athwal, B. S., Friston, K. J., Frackowiak, R. S., & Holmes, A. P. (2000). Variability in fMRI: an examination of intersession differences. *Neuroimage*, 11(6), 708–734.
- Michel*, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, Pickett, K. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden*, E. L. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, 331. Published online ahead of print: 12/16/2010.
- Miller, M. B. & Van Horn, J. D. (2007). Individual variability in brain activations associated with episodic retrieval: a role for large-scale databases. *International Journal of Psychophysiology*, 63(2), 205–213.
- Miller, M. B., Van Horn, J. D., Wolford, G. L., Handy, T. C., Valsangkar-Smyth, M., Inati, S., Grafton, S. T., & Gazzaniga, M. S. (2002). Extensive individual differences in brain activations associated with episodic retrieval are reliable over time. *Journal of Cognitive Neuroscience*, 14(8), 1200–1214.
- Millikan, R. G. (2000). *On Clear and Confused Ideas: An Essay About Substance Concepts*. New York: Cambridge University Press.
- Narayanan, S. (1997). *KARMA: Knowledge-based Action Representations for Metaphor and Aspect*. PhD thesis, University of California at Berkeley.
- Osherson, D. N. & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9, 35–58.
- Osherson, D. N. & Smith, E. E. (1988). Conceptual combination with prototype concepts. In A. Collins & E. Smith (Eds.), *Readings in cognitive science*. Morgan Kaufman, San Mateo CA.
- Peleg, O. & Giora, R. (2011). Salient meanings: The whens and wheres. In K. M. Jaszcolt & K. Allan (Eds.), *Salience and Defaults in Utterance Processing. Mouton Series in Pragmatics* (pp. 35–52). Mouton de Gruyter, New York. General Editor: Istvan Kecskes.
- Peleg, O., Giora, R., & Fein, O. (2004). Contextual strength: The whens and hows of context effects. In I. Noveck & D. Sperber (Eds.), *Experimental Pragmatics* (pp. 172–186). Basingstoke: Palgrave.

- Peleg, O., Giora, R., & Fein, O. (2008). Resisting contextual information: You can't put a salient meaning down. *Lodz Papers in Pragmatics*, 4(1), 13–44.
- Penrose, L. & Penrose, R. (1958). Impossible objects: A special type of visual illusion. *British Journal of Psychology*, 49, 31–33.
- Perrone-Bertolotti, M., Kujala, J., Vidal, J. R., Hamame, C. M., Ossandon, T., Bertrand, O., Minotti, L., Kahane, P., Jerbi, K., & Lachaux, J.-P. (2012). How silent is silent reading? Intracerebral evidence for top-down activation of temporal voice areas during reading. *Journal of Neuroscience*, 32(49), 17554–17562.
- Perrone-Bertolotti, M., Rapin, L., Lachaux, J.-P., Baciú, M., & Løevenbruck, H. (2014). What is that little voice inside my head? inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research*, 261, 220–239.
- Pilkington, A. (2000). *Poetic Effects: A Relevance Theory Perspective*. Amsterdam: John Benjamins.
- Princeton University (2010). About WordNet. WordNet available online at <http://wordnet.princeton.edu>.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: <http://www.R-project.org/>.
- Rayner, K., White, S. J., Johnson, R. L., & Liversedge, S. (2006). Reading Words With Jumbled Letters: There Is a Cost. *Psychological Science*, 17(3), 192–193.
- Reddy, M. J. (1979). The Conduit Metaphor: A Case of Frame Conflict in our Language about Language. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 164–201). Cambridge University Press, Cambridge.
- Regier, T. (1991). Learning Spatial Concepts Using a Partially-Structured Connectionist Architecture. TR-91-050. International Computer Science Institute, Berkeley.
- Regier, T. & Carlson, L. A. (2001). Grounding Spatial Language in Perception: An Empirical And Computational Investigation. *Journal of Experimental Psychology: General*, 130(2), 273–298.
- Rosch, E. (1978). Principles of Categorization. In E. Tosch & B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27–48). Lawrence Erlbaum Associates, Hillsdale NJ.
- Rosch, E. & Mervis, C. B. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, 7, 573–605.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Bream, P. (1976). Basic Objects in Natural Categories. *Cognitive Psychology*, 8, 382–439.
- Rosch, E. H. (1973). Natural Categories. *Cognitive Psychology*, 4, 328–350.
- Rosch, E. H. (1975). Cognitive Reference Points. *Cognitive Psychology*, 7, 532–547.

- Rosch, E. H. (1983). Prototype Classification and Logical Classification: The Two Systems. In E. K. Scholnick (Ed.), *New Trends in Conceptual Representation: Challenges to Piaget's Theory?* (pp. 73–86). Lawrence Erlbaum Associates, Hillsdale.
- Sapir, E. (1921). *Language: An Introduction to the Study of Speech*. Brace: Harcourt.
- Schielzeth, H. & Forstmeier, W. (2009). Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology*, *20*, 416–420.
- Searle, J. R. (1969). *Speech Acts*. New York: Cambridge University Press.
- Searle, J. R. (1975). Indirect speech acts. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics 3: Speech Acts* (pp. 59–82). Academic Press, New York.
- Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge University.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423, 623–656.
- Shannon, C. E. & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana, Illinois: The University of Illinois Press.
- Silver, M. A. & Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends in Cognitive Science*, *13*(11), 488–495.
- Smith, E. E. & Osherson, D. N. (1984). Conceptual Combination with Prototype Concepts. *Cognitive Science*, *8*, 337–361.
- Sperber, D. & Wilson, D. (1986). *Relevance: Communication and Cognition*. Oxford: Blackwell.
- Sperber, D. & Wilson, D. (1995). *Relevance: Communication and Cognition* (2nd ed.). Oxford: Blackwell.
- Sperber, D. & Wilson, D. (1998). The mapping between the mental and the public lexicon. In P. Carruthers & J. Boucher (Eds.), *Thought and Language* (pp. 184–200). Cambridge University Press, Cambridge.
- Staub, A. & Rayner, K. (2007). Eye movements and on-line comprehension processes. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 327–342). Oxford, UK: Oxford University Press.
- Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: Evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 1162–1169.
- Storms, G., De Boeck, P., & Ruts, W. (2000). Prototype and Exemplar-Based Information in Natural Language Categories. *Journal of Memory and Language*, *42*, 51–73.
- Talmy, L. (1983). How Language Structures Space. Technical Report 4, Institute of Cognitive Studies, University of California at Berkeley, January 1983.
- Thibodeau, P. H. & Durgin, F. H. (2011). Metaphor aptness and conventionality: A processing fluency account. *Metaphor and Symbol*, *26*(3), 206–226.

- Thompson-Schill, S. L., Braver, T. S., & Jonides, J. (2005). Individual differences. *Cognitive, Affect, & Behavioral Neuroscience*, 5(2), 115–116.
- Tversky, A. (2004). *Preference, Belief, and Similarity*. Cambridge MA: The MIT Press.
- UNESCO (2013). Index Translationum. Available online at <http://www.unesco.org/xtrans/>.
- University of Oxford (2007). The British National Corpus. Version 3 (BNC XML Edition). Distributed by Oxford University Computing Services on behalf of the the BNC Consortium. URL: <http://www.natcorp.ox.ac.uk/>.
- van der Henst, J.-B., Carles, L., & Sperber, D. (2002). Truthfulness and relevance in telling the time. *Mind & Language*, 17, 457–466.
- van der Henst, J.-B. & Sperber, D. (2004). Testing the cognitive and communicative principles of relevance. In I. Noveck & D. Sperber (Eds.), *Experimental Pragmatics* (pp. 141–171). Palgrave Macmillan, London.
- Van Horn, J. D., Grafton, S. T., & Miller, M. B. (2008). Individual variability in brain activity: A nuisance or an opportunity? *Brain Imaging Behavior*, 2(4), 327–334.
- van Rossum, G. & Python Software Foundation (1991). Python programming language. URL: <http://www.python.org/>.
- Whorf, B. L. (1940). Language and Science. *MIT Technology Review*, 35, 229–231, 247–248.
- Whorf, B. L. (1956). *Language, Thought, and Reality*. Cambridge MA: MIT Press.
- Wierzbicka, A. (1992). Defining Emotion Concepts. *Cognitive Science*, 16, 539–581.
- Wilks, S. S. (1938). The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, 9(1), 60–62.
- Wilson, D. & Sperber, D. (1993). Linguistic form and relevance. *Lingua*, 90(1/2), 1–25.
- Wilson, D. & Sperber, D. (2002). Truthfulness and relevance. *Mind*, 111, 583–632.
- Wilson, D. & Sperber, D. (2004). Relevance Theory. In L. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 607–632). Blackwell, Oxford.
- Wittgenstein, L. (1953). *Philosophische Untersuchungen (Philosophical Investigations)*. Oxford: Basil Blackwell. Translated by G. E. M. Anscombe.
- Wittgenstein, L. (1958). *Philosophische Untersuchungen (Philosophical Investigations)* (2nd ed.). Oxford: Basil Blackwell. Translated by G. E. M. Anscombe.
- Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine*, 22(22), 3527–3541.
- Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex no access. *Journal of Cognitive Neuroscience*, 23(10), 3146–3152.
- Zatorre, R. J. & Halpern, A. R. (2005). Mental concerts: Musical imagery and auditory cortex. *Neuron*, 47, 9–12.

Appendix A

A.1 Experimental material

Table A.1: Metaphors used in the experiments, English and German version. Contextual continuations in brackets.

English
<i>English-specific metaphors</i>
Talk is cheap (in its exuberance)
Love is a journey (with obstacles)
A smile is a magnet (for friendly people)
Words are daggers (out of your mouth)
Love is a battlefield (of love-rivals)
Revenge is a bitch (in its baseness)
A home is a castle (to retreat to)
Truth is stranger than fiction (in its unimaginable possibilities)
Beauty is skin-deep (in its superficiality)
Custom is a second nature (in its influence on us)
Praise is not pudding (if it is hollow)
Promise is debt (until it is kept)
<i>German-specific metaphors</i>
Dreams are froth (on the ocean of wishes)
Art is a science (of experimenting)
Beauty is a ticket (to fame)
Election day is pay day (for politicians)
Faith is a fortress (in its perseverance)
Time is an ocean (in a storm)
Sports is murder (for the non-athletic)
The devil is a squirrel (with malice)
Following is suffering (in the eyes of the prosecution)
Laughter is a bridge (between people)
The spine is a pillar (of bones)
Hope is patience (in the face of adversities)
<i>Cross-linguistically shared metaphors</i>
Life is a journey (on the road of uncertainty)
Time is money (in the working world)
Seeing is believing (in its reliability)
Love is blind (in the beginning)
Ignorance is bliss (for the unknowing)
Silence is golden (in a solemn moment)
Jealousy is poison (in a relationship)
Revenge is sweet (with satisfaction)

The mind is a sponge (to absorb knowledge)
The brain is a computer (with the mind as software)
A job is a jail (with its restrictions)
Patience is a virtue (in all areas of life)
Knowledge is power (over those without it)
Anger is a volcano (of emotion)

Novel metaphors

Love is a castle (of fairy tales)
A home is a sponge (of comfort)
Beauty is a computer (of calculability)
Time is a squirrel (with unpredictable jumps)
Silence is a volcano (under pressure)
A friend is a magnet (for happiness)
Love is murder (for selfishness)
A home is money (for the future)
Beauty is patience (on the way to perfection)
Time is believing (with continuity)
Silence is blind (until the end of silence)
A friend is cheap (until you value them)

German

English-specific metaphors

Reden ist billig (in seiner Überschwänglichkeit)
Liebe ist eine Reise (mit Hindernissen)
Ein Lächeln ist ein Magnet (für freundliche Menschen)
Worte sind Dolche (aus deinem Munde)
Die Liebe ist ein Schlachtfeld (der Nebenbuhler)
Rache ist eine Zicke (in ihrer Niederträchtigkeit)
Ein Zuhause ist eine Burg (in die man sich zurückziehen kann)
Die Wahrheit ist komischer als die Fiktion (in ihren ungeahnten Möglichkeiten)
Schönheit ist hauchdünn (in ihrer Oberflächlichkeit)
Gewohnheit ist eine zweite Haut (in ihrem Einfluss auf uns)
Lob ist kein Pudding (wenn es hohl ist)
Ein Versprechen ist eine Schuld (bis es eingelöst wird)

German-specific metaphors

Träume sind Schäume (auf dem Ozean der Wünsche)
Kunst ist eine Wissenschaft (des Experimentierens)
Schönheit ist eine Eintrittskarte (zum Ruhm)
Wahltag ist Zahltag (für Politiker)
Der Glaube ist eine Festung (in seiner Beharrlichkeit)
Zeit ist ein Ozean (im Sturm)
Sport ist Mord (für unsportliche Leute)
Der Teufel ist ein Eichhörnchen (mit bösen Absichten)
Mitgegangen ist Mitgefangen (in den Augen des Anklägers)
Lachen ist eine Brücke (zwischen Menschen)
Das Rückgrat ist eine Säule (aus Knochen)
Hoffnung ist Geduld (angesichts von Widrigkeiten)

Cross-linguistically shared metaphors

Das Leben ist eine Reise (auf der Straße der Ungewissheit)
Zeit ist Geld (in der Arbeitswelt)
Sehen ist Glauben (in seiner Verlässlichkeit)
Liebe ist blind (am Anfang)
Unwissenheit ist ein Segen (für den Unwissenden)
Schweigen ist Gold (in Momenten der Andacht)
Eifersucht ist Gift (in einer Beziehung)
Rache ist süß (mit Genugtuung)
Der Verstand ist ein Schwamm (der Wissen aufsaugt)

Das Gehirn ist ein Computer (mit dem Verstand als Software)
 Ein Beruf ist ein Gefängnis (mit seinen Beschränkungen)
 Geduld ist eine Tugend (in allen Lebenslagen)
 Wissen ist Macht (über die Unwissenden)
 Wut ist ein Vulkan (von Emotionen)
Novel metaphors
 Liebe ist eine Burg (aus dem Märchen)
 Ein Zuhause ist ein Schwamm (des Wohlbehagens)
 Schönheit ist ein Computer (der Berechenbarkeit)
 Zeit ist ein Eichhörnchen (mit unberechenbaren Sprüngen)
 Schweigen ist ein Vulkan (unter Druck)
 Ein Freund ist ein Magnet (für Freude)
 Liebe ist Mord (für Selbstsüchtigkeit)
 Ein Zuhause ist Geld (für die Zukunft)
 Schönheit ist Geduld (auf dem Weg zur Perfektion)
 Zeit ist Glauben (mit Stetigkeit)
 Schweigen ist blind (bis zum Ende der Stille)
 Ein Freund ist billig (bis man ihn schätzt)

A.2 Corpus frequencies

Table A.2: Word frequencies (tokens per million words in the corpus). British National Corpus (BNC), Google n-gram corpus of books published in the UK (G-UK), Corpus of Contemporary American English (COCA), Google n-gram corpus of books published in the US (G-US), corpus of written contemporary German (C-W) from the COSMAS-2 project, and the Google n-gram corpus of books published in Germany (G-GE). Frequency is the number of tokens per one million word in the corpus. Columns p_E and p_G give the cumulative probability according to the distribution in Figure (5.3). Sorted by descending frequency and corrected for corpus size.

	BNC	G-UK	COCA	G-US	p_E	C-W	G-GE	p_G
Time	1868.4	1295.6	1934.3	1369.5	.741	780.2	923.6	.641
Life	565.3	533.8	689.3	562.7	.642	515.3	424.1	.587
Home	579.0	257.9	665.0	319.2	.570	23.6	2.9	.499
Words	433.0	348.9	382.7	363.6	.537	125.7	180.7	.501
Following	270.7	271.8	148.8	277.5	.525	0.1	0.0	.319
Mind	307.9	171.6	243.7	179.3	.524	13.3	53.0	.499
Love	228.1	149.9	330.3	194.0	.524	118.5	163.6	.501
Money	374.3	142.0	406.2	160.8	.524	304.3	80.1	.506
Friend	317.5	153.7	358.8	196.6	.522	193.8	56.4	.500
Job	325.7	141.7	391.9	183.1	.521	98.0	41.4	.499
Art	210.3	142.1	287.5	133.7	.510	138.0	187.5	.502
Hope	218.8	96.8	217.9	99.4	.504	96.2	61.4	.499
Talk	166.4	97.9	322.6	120.6	.501	136.0	137.7	.501
Computer	176.2	97.0	160.7	137.2	.500	53.5	24.9	.499
Science	129.7	131.2	169.9	106.3	.500	33.1	167.8	.500
Sea	138.4	89.5	100.4	68.0	.500	34.5	36.0	.499
Faith	53.6	62.4	72.4	65.6	.500	69.3	89.8	.499
Brain	56.9	74.7	85.0	90.3	.500	12.6	20.3	.499
Seeing	67.4	42.2	95.7	42.7	.500	309.8	17.2	.502
Smile	79.9	23.6	101.1	76.2	.500	19.2	7.3	.499
Bridge	75.7	45.6	61.1	36.5	.500	58.4	22.7	.499

Dreams	70.9	53.6	109.1	67.1	.500	73.2	25.7	.499
Sport(s)	90.6	37.6	146.8	37.8	.500	127.0	23.2	.499
Beauty	44.1	33.4	53.0	34.1	.500	21.4	47.8	.499
Trip	59.4	33.0	108.3	47.4	.500	4.1	0.1	.498
Silence	57.2	30.8	54.9	36.0	.500	20.7	13.8	.499
Sweet	35.6	21.4	60.2	26.3	.500	22.5	1.8	.499
Journey	56.2	39.7	35.6	28.5	.500	125.9	51.6	.499
Virtue	25.8	45.2	21.5	33.8	.500	7.4	42.2	.499
Suffering	42.8	45.6	37.7	40.7	.500	0.1	0.0	.304
Anger	37.8	26.8	41.4	36.8	.500	12.3	8.7	.499
Castle	61.9	20.4	17.1	10.2	.500	51.6	12.9	.499
Golden	38.4	16.6	50.4	17.1	.500	52.4	32.1	.499
Blind	31.4	23.9	32.9	24.4	.500	19.3	8.8	.499
Ocean	25.4	15.4	49.6	22.2	.500	5.7	5.6	.499
Ticket	47.3	15.3	59.7	17.0	.500	18.2	0.4	.499
Cheap	39.0	20.2	29.3	12.9	.500	41.6	14.2	.499
Murder	7.9	22.3	11.7	27.9	.500	41.3	10.9	.499
Devil	19.6	10.1	21.2	11.0	.500	20.3	15.6	.499
Patience	11.9	9.2	16.1	10.7	.500	17.9	11.3	.499
Ignorance	11.6	13.7	9.1	11.1	.500	1.5	9.6	.499
Laughter	21.2	10.4	28.7	13.2	.500	32.8	9.6	.499
Jail	13.4	5.4	39.9	13.9	.500	29.5	14.0	.499
Believing	14.0	11.5	12.0	10.5	.500	46.4	45.2	.499
Spine	14.8	16.0	12.7	18.1	.500	2.8	2.0	.498
Poison	11.6	7.3	11.9	9.3	.499	7.9	7.7	.499
Revenge	10.7	8.1	10.7	8.0	.499	11.1	12.9	.499
Jealousy	7.7	6.2	5.3	5.9	.499	4.3	6.9	.499
Pillar	9.9	8.9	7.7	6.3	.499	11.9	10.0	.499
Fortress	6.3	7.1	4.5	4.5	.499	12.0	14.3	.499
Magnet	4.7	4.7	8.7	5.4	.499	3.6	3.3	.499
Bliss	3.9	3.0	4.8	2.8	.498	12.0	14.0	.499
Battlefield	4.0	5.6	9.8	5.3	.498	2.2	2.9	.498
Volcano	7.1	3.8	8.6	3.5	.497	7.2	1.8	.499
Sponge	6.1	3.4	5.2	4.5	.497	5.6	1.4	.499
Dagger	3.9	2.7	3.5	2.8	.496	0.8	1.9	.496
Squirrel	4.0	1.9	7.2	3.2	.496	1.7	0.6	.496
Election day	0.4	0.3	4.1	0.8	.487	3.5	0.9	.498
Froth	1.4	0.7	1.0	0.7	.481	2.2	1.8	.498
Pay day	0.7	0.2	2.0	0.5	.466	0.8	0.1	.482

Table A.3: Metaphor frequencies (token counts). British National Corpus (BNC), Google n-gram corpus of books published in the UK (G-UK), Corpus of Contemporary American English (COCA), Corpus of Historical American English (COHA), Google n-gram corpus of books published in the US (G-US) and fiction books (G-FIC), corpus of written contemporary German (C-W) and of historical German (CH) from the COSMAS-2 project, and the Google n-gram corpus of books published in Germany (G-GE). None of the novel metaphor were found in any of the 6 English and 3 German corpora. Sorted by metaphor type.

English corpora	BNC	G-UK	COCA	COHA	G-US	G-FIC
<i>English-specific</i>						
Words are daggers	0	42	0	3	50	46
Love is a journey	0	0	1	0	109	42
Love is a battlefield	0	0	3	0	127	0
Seeing is believing	4	3,649	62	39	13,347	7,355
A home is a castle	6	634	12	15	3,016	NA
Talk is cheap	0	253	60	24	3,488	2,235
Ignorance is bliss	4	2,311	35	51	7,271	5,057
The brain is a computer	0	45	0	0	146	60
Revenge is a bitch	0	0	0	0	0	0
<i>German-specific</i>						
Following is suffering	0	0	0	0	0	0
Art is a science	0	0	0	0	85	0
Sport is murder	0	0	0	0	0	0
Election day is pay day	0	0	0	0	0	0
Dreams are froth	0	0	0	0	0	0
The devil is a squirrel	0	0	0	0	0	0
<i>Shared</i>						
Knowledge is power	15	7,057	117	95	26,079	12,617
Patience is a virtue	4	723	41	11	2,636	2,248
Time is money	22	3,694	88	69	17,935	10,164
Life is a trip/journey	1	523	13	5	1,948	1,185
Revenge is sweet	2	754	9	18	2,111	1,949
Silence is golden	4	1,281	26	27	5,224	3,269
Love is blind	3	2,508	28	50	7,856	5,763
German corpora	C-W	C-H	G-GE			
<i>English-specific</i>						
Words are daggers	0	0	0			
Love is a journey	0	0	0			
Love is a battlefield	0	0	0			
Seeing is believing	0	0	0			
A home is a castle	0	0	0			
Talk is cheap	0	0	0			
Ignorance is bliss	0	0	0			
The brain is a computer	0	0	0			
Revenge is a bitch	0	0	0			
<i>German-specific</i>						
Following is suffering	2	0	0			
Art is a science	1	0	0			
Sport is murder	258	0	184			
Election day is pay day	181	0	145			
Dreams are froth	86	14	618			
The devil is a squirrel	14	0	0			
<i>Shared</i>						
Knowledge is power	323	6	7,637			

Patience is a virtue	23	0	48
Time is money	702	3	3,129
Life is a trip/journey	26	0	0
Revenge is sweet	284	5	195
Silence is golden	300	5	1,003
Love is blind	3	1	164

Table A.4: Diachronic change in metaphors' corpus frequency. Present-day frequency f_m , and the frequency f_{30} 30 years ago. Frequencies are in tokens per million words. In the case of culturally-shared metaphors, English and German corpora are treated separately to identify discrepancies of f_m and α_m .

	Metaphor	f_m	f_{30}
English	Love is a battlefield	.00238	.00213
	Love is a journey	.00163	.00029
	Talk is cheap	.03889	.01894
	A home is a castle	.10760	.02938
	Words are daggers	.00025	.00004
Misc	Truth is stranger than fiction	.02251	.01902
	Promise is debt	.00033	.00082
	Beauty is skin-deep	.00115	.00112
	Custom is a second nature	.00089	.00055
German	Sport is murder	.01880	.00580
	Election day is pay day	.00676	.00657
	Dreams are froth	.02177	.01340
Shared (E corpora)	Knowledge is power	.24140	.12469
	Patience is a virtue	.03364	.01691
	Time is money	.17908	.12206
	Life is a journey	.02777	.01050
	The brain is a computer	.00271	.00148
	Seeing is believing	.12879	.09693
	Ignorance is bliss	.05614	.03624
	Revenge is sweet	.01547	.01141
	Silence is golden	.03998	.02947
	Silence is gold	.00170	.00208
	Love is blind	.05372	.04555
	Shared (G corpora)	Knowledge is power	.30471
Patience is a virtue		.00298	.00051
Time is money		.16581	.07053
Life is a journey		0	0
Revenge is sweet		.01082	.00329
Silence is golden		0	0
Silence is gold		.05734	.03033
Love is blind		.00338	.00329

English-specific metaphors over time

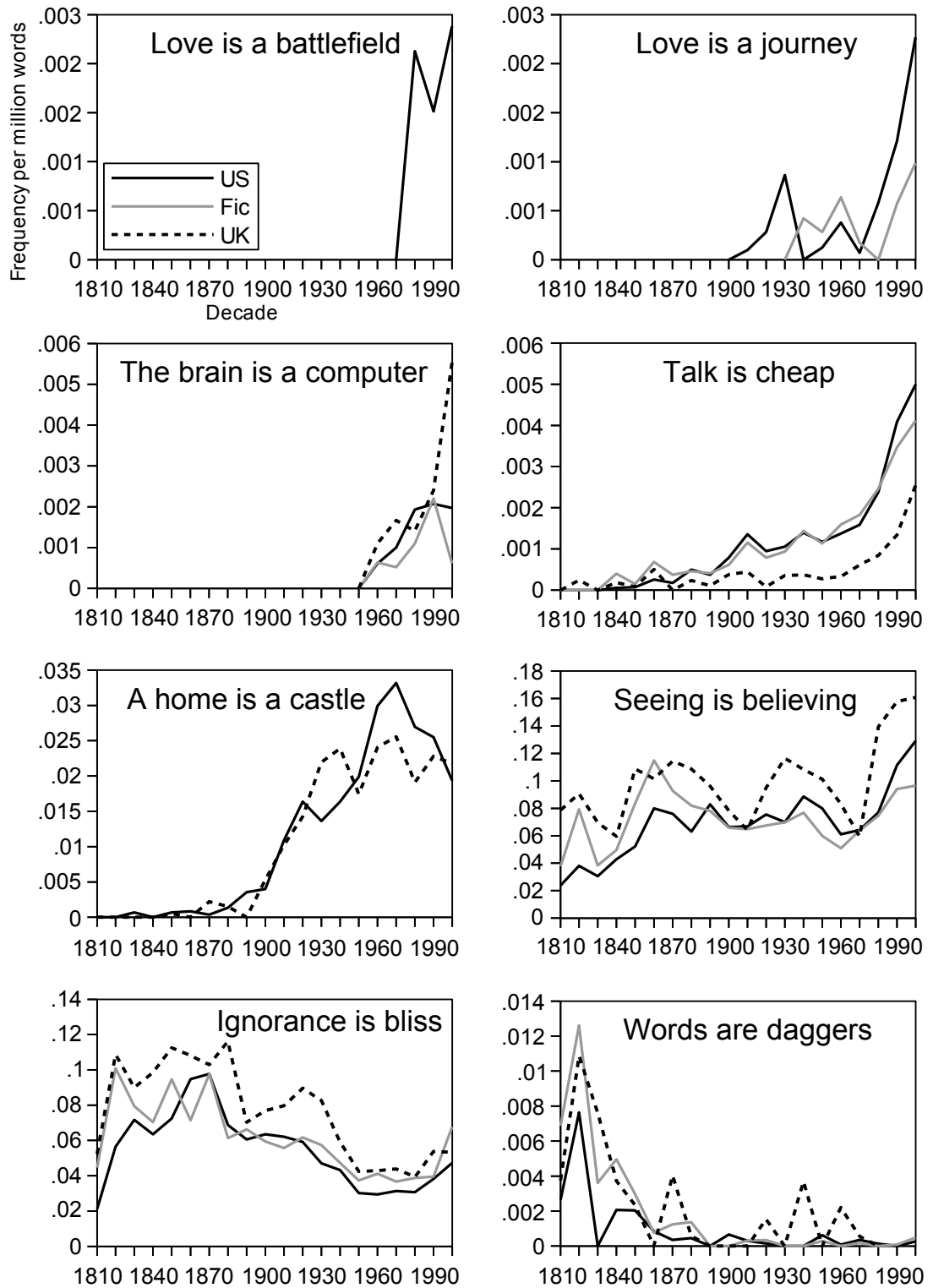
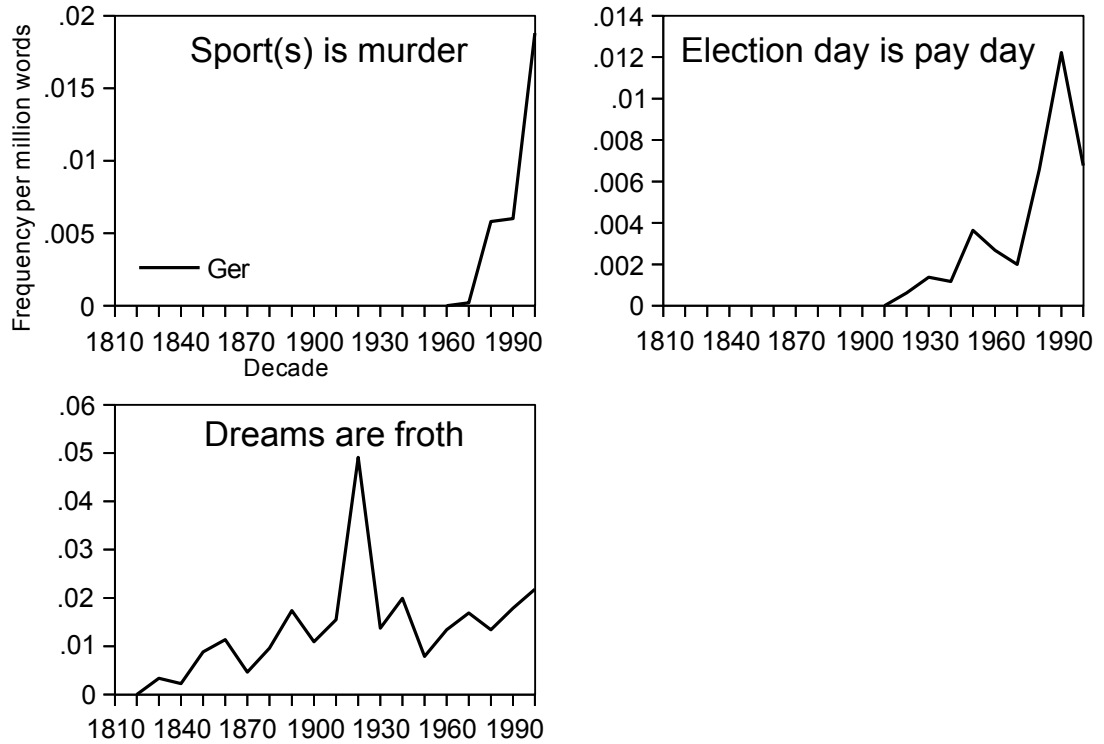


Figure A.1: Diachronic change in frequency: English-specific metaphors.

German-specific metaphors over time



Other metaphoric idioms

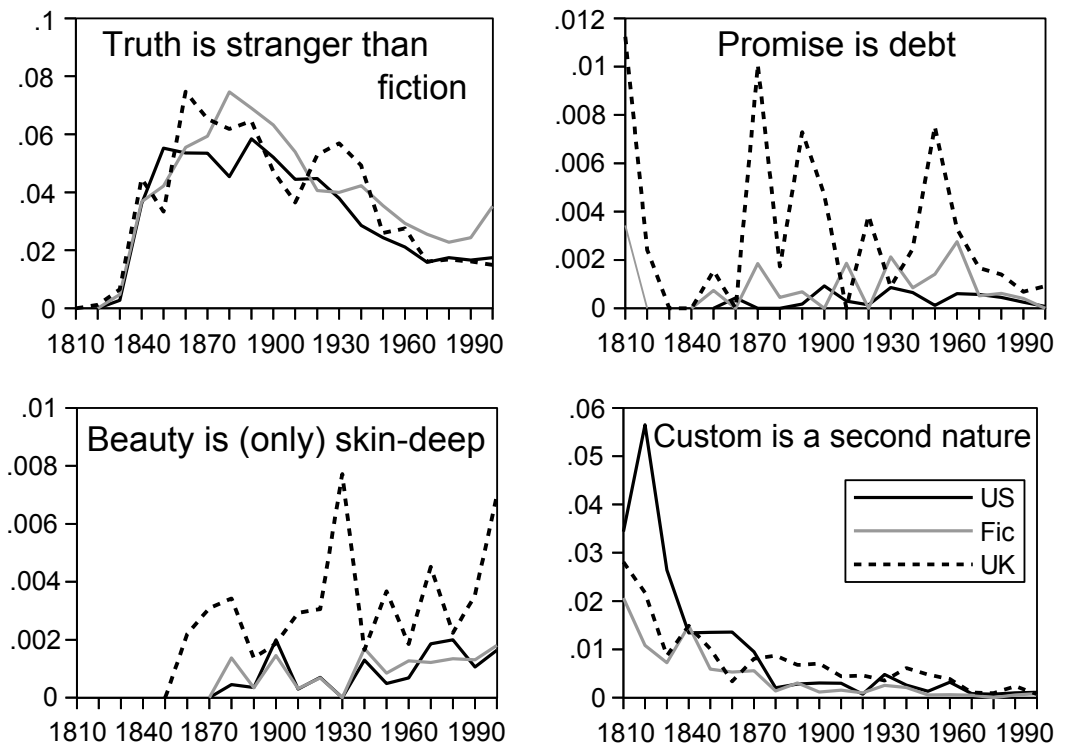


Figure A.2: Diachronic change in frequency: German-specific metaphors.

Cross-linguistically shared metaphors over time

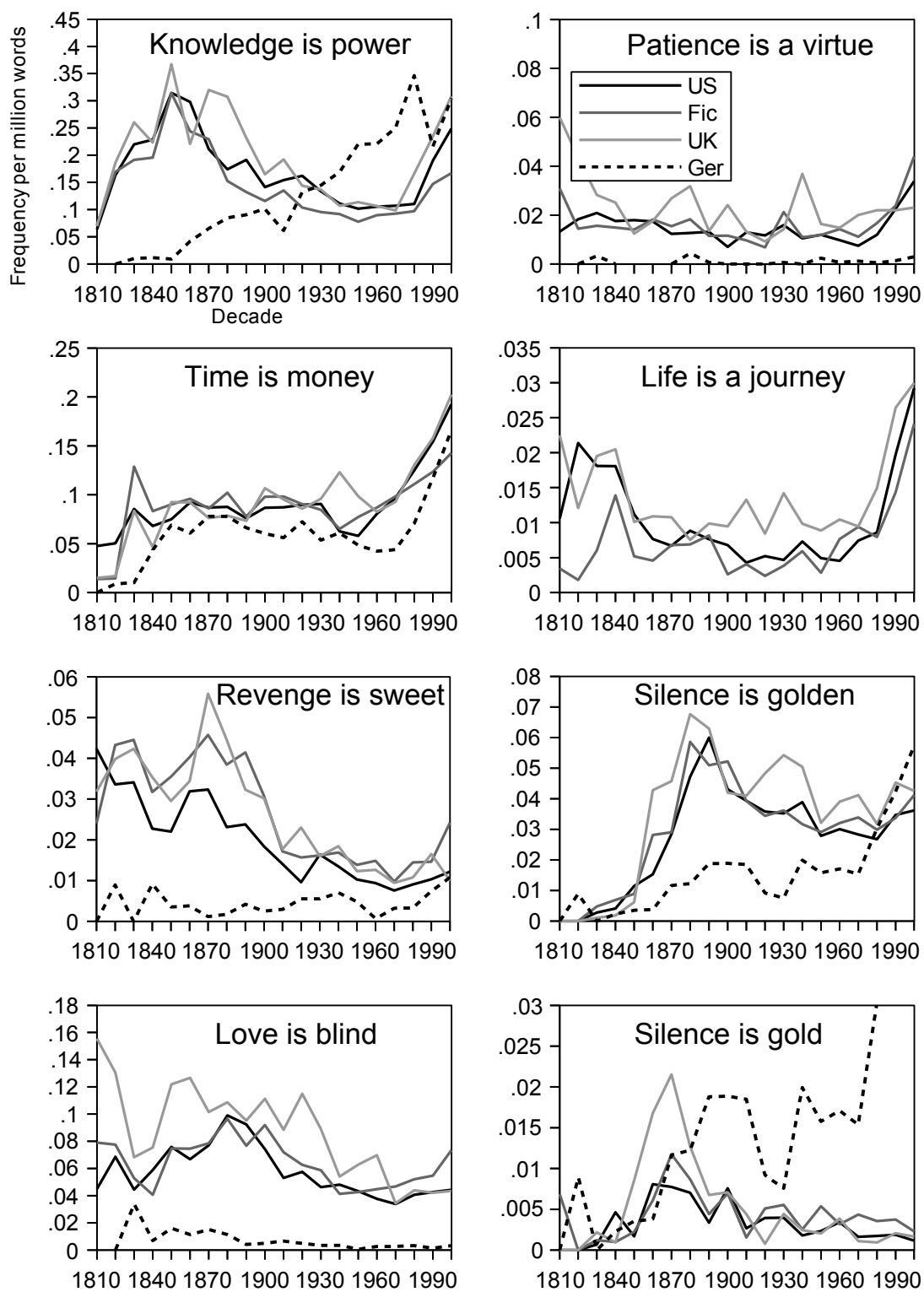


Figure A.3: Diachronic change in frequency: culturally-shared metaphors.

A.3 Semantic Levenshtein distances

Table A.5: Metaphor type, relative semantic Levenshtein distance, relevant and irrelevant sense differences for all translation pairs. Note, words can appear in multiple metaphor types (E = English-specific, G = German-specific, S = shared, N = novel).

English	German	Type	Dist.	Irrel.	Rel.
Anger	Wut	S	.25	1	0
Art	Kunst	G	.25	1	0
Battlefield	Schlachtfeld	E	0	NA	NA
Beauty	Schönheit	G, E, N	0	NA	NA
Believing	Glaube(n)	S, N	0	NA	NA
Blind	blind	S, N	0	NA	NA
Bliss	Segen	S	.333	0	1
Brain	Gehirn	S	.2	NA	NA
Bridge	Brücke	G	.462	6	0
Castle	Burg	E, N	.5	1	0
Cheap	billig	E, N	0	NA	NA
Computer	Computer	S, N	0	NA	NA
Dagger	Dolch	E	0	NA	NA
Devil	Teufel	G	0	NA	NA
Dreams	Traum	G	0	NA	NA
Election day	Wahltag	G	0	NA	NA
Faith	Glaube(n)	G, N	0	NA	NA
Following	Mitgefangen	G	.5	0	1
Fortress	Festung	G	0	NA	NA
Friend	Freund	N	0	NA	NA
Froth	Schaum	G	.667	1	1
Golden	Gold	S	0	NA	NA
Home	Zuhause	N	.333	3	0
Hope	Hoffnung	G	0	NA	NA
Ignorance	Unwissenheit	S	0	NA	NA
Jail	Gefängnis	S	0	NA	NA
Jealousy	Eifersucht	S	.5	1	0
Job	Beruf	S	.667	1	1
Journey	Reise	G	0	NA	NA
Knowledge	Wissen	S	0	NA	NA
Laughter	Lachen	G	0	NA	NA
Life	Leben	S	0	NA	NA
Love (N)	Liebe	E, S, N	.167	1	0
Love (V)	Liebe	E, S, N	.167	1	0
Magnet	Magnet	E	0	NA	NA
Mind	Verstand	S	.429	3	0
Money	Geld	S, N	0	NA	NA
Murder	Mord	G, N	0	NA	NA
Ocean	Ozean	G	0	NA	NA
Patience	Geduld	G, S, N	0	NA	NA
Pay day	Zahltag	G	0	NA	NA
Pillar	Säule	G	.2	1	0
Poison	Gift	S	0	NA	NA
Power	Macht	S	0	NA	NA
Revenge	Rache	E, S	0	NA	NA
Science	Wissenschaft	G	0	NA	NA
Seeing	Sehen	S	0	NA	NA
Silence	Schweigen	S, N	0	NA	NA

Smile	Lächeln	E	0	NA	NA
Spine	Rückgrat	G	.75	3	0
Sponge	Schwamm	S, N	.5	2	0
Sport(s)	Sport	G	.4	2	0
Squirrel	Eichhörnchen	G, N	.5	1	0
Suffering	Mitgehangen	G	.75	3	0
Sweet	süß	S	0	NA	NA
Talk	Rede(n)	E	0	NA	NA
Ticket	Eintrittskarte	G	.75	2	1
Time	Zeit	G, S, N	.4	4	0
Trip	Trip	E	.5	3	0
Virtue	Tugend	S	0	NA	NA
Volcano	Vulkan	S	0	NA	NA
Words	Worte	E	.4	1	1

A.4 Statistical analysis of the experimental results

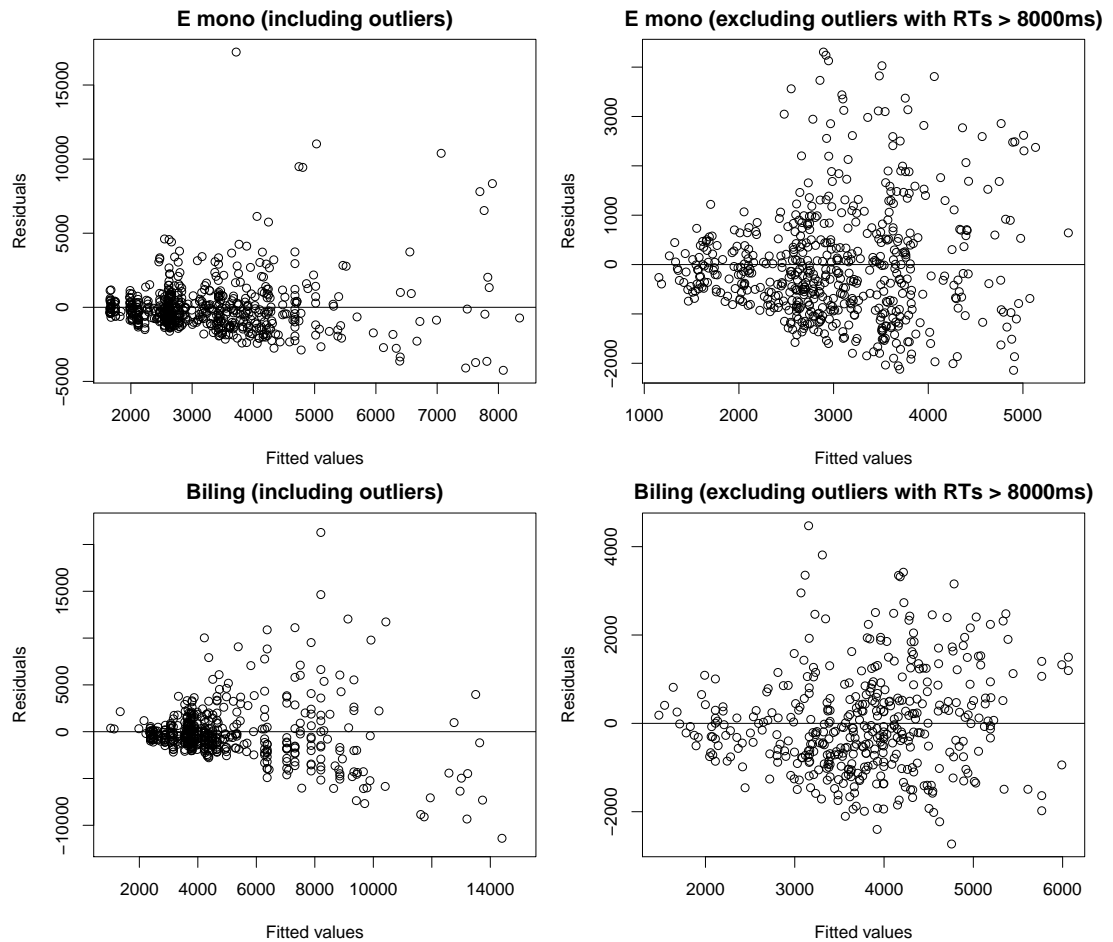


Figure A.4: Residual plots for English monolinguals’ and English-German bilinguals’ reading/response times (RT). For both English monolinguals and bilinguals (see left column of graphs), the variance in RT is not homoscedastic; higher fitted values have larger residuals than smaller values; the (full) models are thus less “good” predictors for larger mean RTs, which, however, is what we would expect as larger RTs correspond to outliers. For the range we are most interested in and to which the mixed models were fitted, RTs faster than 8000 milliseconds, the variance *is* homoscedastic (see the graphs in the right column).

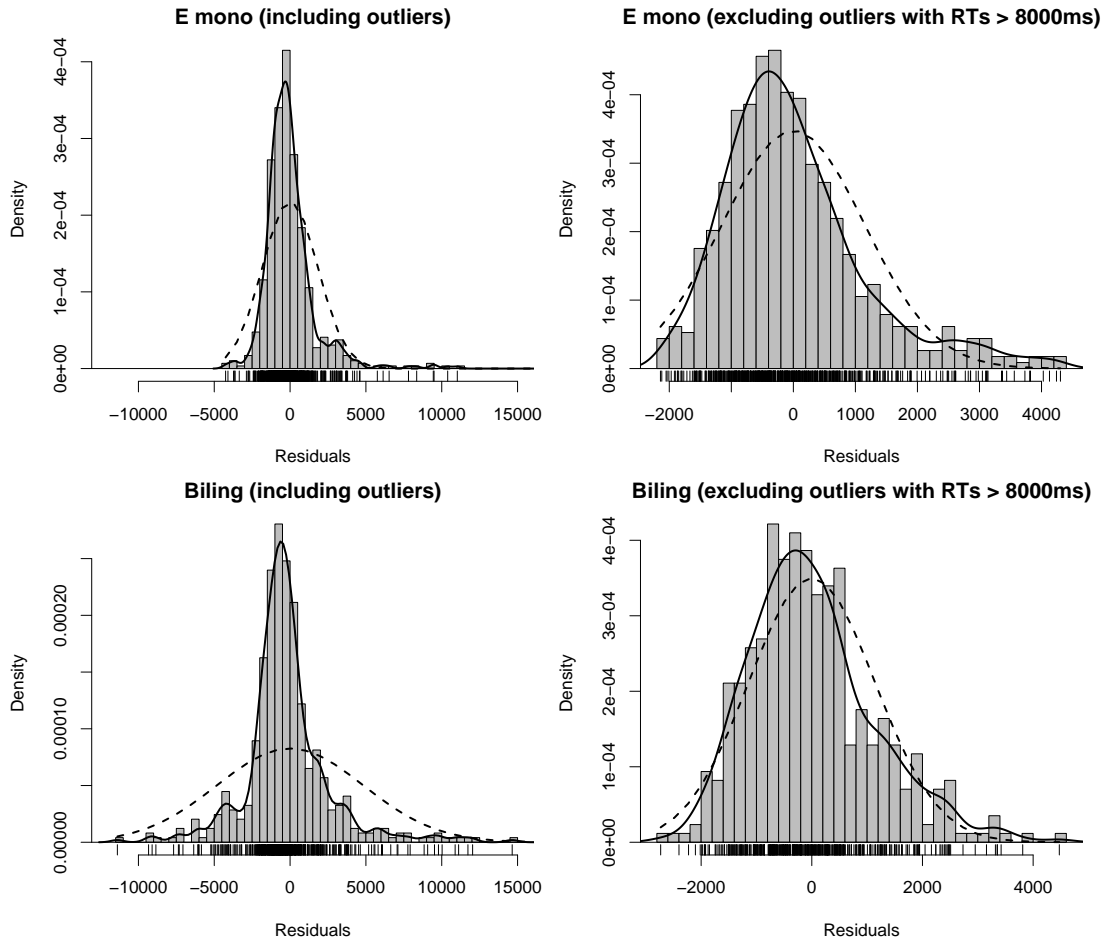


Figure A.5: The histograms give the probability densities for English monolinguals and English-German bilinguals. The histograms in the left column are the raw residuals; in the right column, RTs over 8000 ms have been excluded. The actual kernel density of RTs is plotted as a continuous line. The corresponding normal distribution is plotted as a dashed line. The actual residuals (as given in Figure A.4) are plotted as a rug along the x-axis. By comparing the kernel densities to the normal densities we can see that in the raw data bilinguals' RTs are not normally distributed, while monolinguals' RTs are. When RTs over 8000 ms are excluded, however, the kernel densities of both monolinguals and bilinguals are normally distributed along the regression of the corresponding linear mixed effects models.

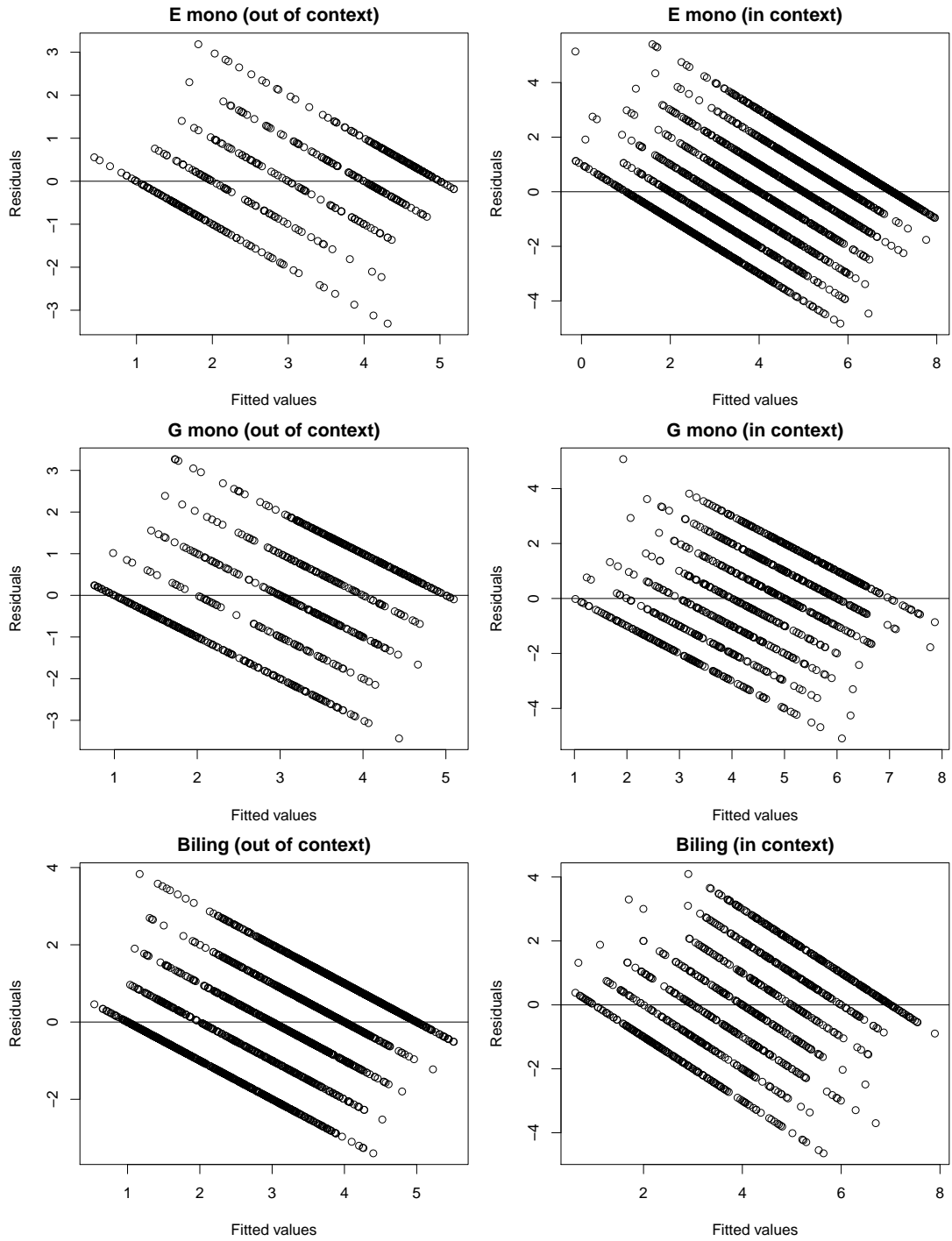


Figure A.6: Residual plots for out-of-context and in-context plausibility judgements. Fitted linear mixed effects models. The stripes are a result of the Likert scale used to record plausibility judgements: a 5-point Likert scale for the out-of-context condition and a 7-point scale for the in-context condition. However, notice that each stripe spans all fitted values and at least half the residuals. This means that although plausibility ratings are restricted to the increments of the respective Likert scale, within that increment, by-subject and by-item variation *is* homoscedastic.

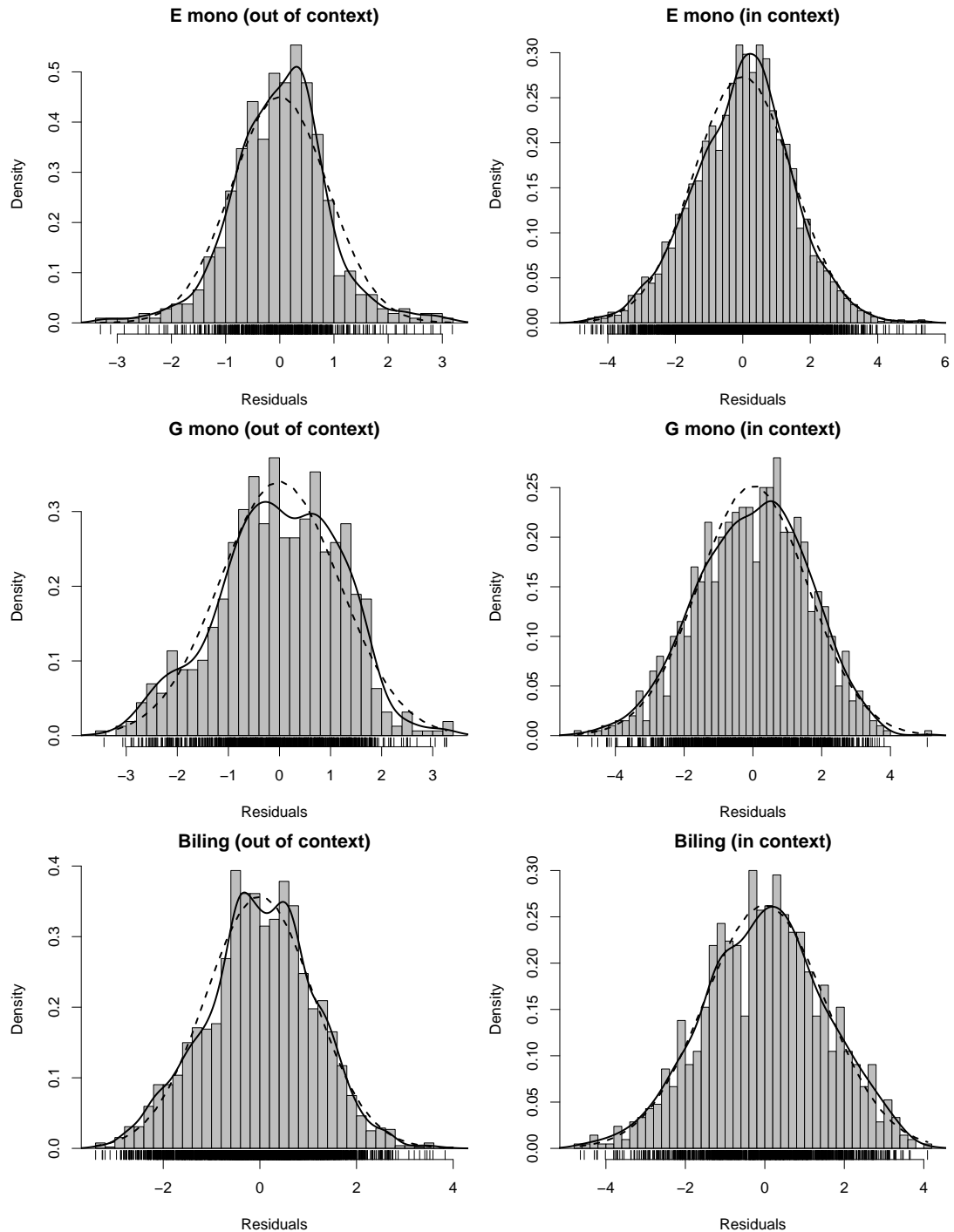


Figure A.7: The histograms give the probability densities for English monolinguals, German monolinguals and English-German bilinguals. The histograms in the left column are the residuals of the out-of-context plausibility judgements; in the right column are the residuals of the in-context plausibility judgements. The actual kernel density of plausibility judgements is plotted as a continuous line. The corresponding normal distribution, i.e. a normal distribution using the mean and standard deviation of the actual distribution, is plotted as a dashed line. The actual residuals (as given in Figure A.6) are plotted as a rug along the x-axis. By comparing the kernel densities to the normal densities we can see that monolinguals' and bilinguals' plausibility judgements are normally distributed. The normality requirement for their linear mixed effects models is therefore satisfied.

A.5 Experimental results and deficits in the material

Table A.6: Cross-linguistic variation in the experimental material. Shown are the number of samples N , the effect size (the difference in cumulative probabilities between English and German), the p -value of statistical significance (** for $p \leq .01$, † for $p \leq .1$), the probability of Type I errors α , the probability of Type II errors β , and statistical power $P = 1 - \beta$ ($\diamond = P < .8$). IDs: (1) Metaphor frequency, (2) diachronic frequency change, (3) word frequency, metaphor length measured as (4) the number of characters, (5) syllables, (6) morphemes, and (7) words, (8) lexical sense differences, (9) contextual relevance of sense differences, and (10) compositional-semantic differences. The last row in each table provides general estimates of central tendency, averaged across the individual studies and weighted according to their sample size.

English-specific metaphors					
ID	N	Effect size	$p = \alpha$	β	P
1	12	.060	.792	.073	.927
2	12	.048	.834	.068	.932
3	12	.005	.983	.052	.948
4	12	.209	.260	.170	.830
5	12	.223	.208	.183	.817
6	12	.089	.690	.088	.912
7	12	.208	.263	.169	.831
8	12	.133	.535	.113	.887
9	12	.140	.510	.118	.882
10	12	.131	.542	.112	.888
	120	.125	.562	.115	.885

German-specific metaphors					
ID	N	Effect size	$p = \alpha$	β	P
1	12	.060	.792	.073	.927
2	12	.018	.938	.056	.944
3	12	.022	.924	.058	.942
4	12	.215	.237	.175	.825
5	12	.275	.009 **	.236	.764 \diamond
6	12	.102	.645	.095	.905
7	12	.125	.564	.108	.892
8	12	.225	.086 †	.214	.786 \diamond
9	12	.110	.617	.099	.901
10	12	.235	.162	.194	.806
	120	.139	.497	.131	.869

Table A.7: Cross-linguistic variation in the experimental material. Shown are the number of samples N , the effect size (the difference in cumulative probabilities between English and German), the p-value of statistical significance, the probability of Type I errors α , the probability of Type II errors β , and statistical power $P = 1 - \beta$. IDs: (1) Metaphor frequency, (2) diachronic frequency change, (3) word frequency, metaphor length measured as (4) the number of characters, (5) syllables, (6) morphemes, and (7) words, (8) lexical sense differences, (9) contextual relevance of sense differences, and (10) compositional-semantic differences. The last row in each table provides general estimates of central tendency, averaged across the individual studies and weighted according to their sample size.

Cross-linguistically shared metaphors					
ID	N	Effect size	$p = \alpha$	β	P
1	14	.045	.831	.069	.931
2	14	.055	.794	.074	.926
3	14	.010	.963	.054	.946
4	14	.158	.401	.140	.860
5	14	.220	.156	.198	.802
6	14	.032	.880	.063	.937
7	14	.071	.734	.082	.918
8	14	.145	.452	.130	.870
9	14	.140	.471	.126	.874
10	14	.152	.425	.135	.865
	140	.103	.611	.107	.893

Novel metaphors					
ID	N	Effect size	$p = \alpha$	β	P
1	12	NA	NA	NA	NA
2	12	NA	NA	NA	NA
3	12	.015	.948	.055	.945
4	12	.177	.377	.144	.856
5	12	.125	.564	.108	.892
6	12	0	1	.050	.950
7	12	0	1	.050	.950
8	12	.143	.499	.120	.880
9	12	0	1	.050	.950
10	12	.107	.627	.097	.903
	120	.071	.752	.084	.916

Table A.8: Cross-linguistic variation in the experimental material, sorted from least to largest variation. Shown are the number of samples N , the effect size (the difference in cumulative probabilities between English and German), the p-value of statistical significance, the probability of Type I errors α , the probability of Type II errors β , and statistical power $P = 1 - \beta$. IDs: (1) Metaphor frequency, (2) diachronic frequency change, (3) word frequency, metaphor length measured as (4) the number of characters, (5) syllables, (6) morphemes, and (7) words, (8) lexical sense differences, (9) contextual relevance of sense differences, and (10) compositional-semantic differences. The last row in each table provides general estimates of central tendency, averaged across the individual studies and weighted according to their sample size.

	N	Effect size	$p = \alpha$	β	P
(3) Word freq	50	.013	.955	.055	.945
(2) Diachronic	50	.041	.852	.066	.934
(1) Meta freq	50	.054	.806	.072	.928
(6) Morphemes	50	.055	.807	.074	.926
(9) Relevance	50	.099	.642	.099	.901
(7) Words	50	.100	.644	.101	.899
(10) Comp sem	50	.156	.438	.135	.865
(8) Lex sem	50	.161	.395	.144	.856
(4) Characters	50	.188	.322	.157	.843
(5) Syllables	50	.211	.231	.182	.818
	500	.108	.609	.108	.892

Table A.9: Cross-linguistic variation in the experimental material, averaged per metaphor type and weighted by sample size. Shown are the number of samples N , the effect size (the difference in cumulative probabilities between English and German), the p-value of statistical significance, the probability of Type I errors α , the probability of Type II errors β , and statistical power $P = 1 - \beta$. IDs: (1) Metaphor frequency, (2) diachronic frequency change, (3) word frequency, metaphor length measured as (4) the number of characters, (5) syllables, (6) morphemes, and (7) words, (8) lexical sense differences, (9) contextual relevance of sense differences, and (10) compositional-semantic differences. The last row in each table provides general estimates of central tendency, averaged across the individual studies and weighted according to their sample size.

Metaphors' linguistic form					
	N	Effect size	$p = \alpha$	β	P
E	84	.120	.576	.115	.885
G	84	.117	.587	.114	.886
S	98	.084	.680	.097	.903
N	84	.063	.778	.081	.919
	350	.096	.656	.102	.898

Metaphors' linguistic meaning					
	N	Effect size	$p = \alpha$	β	P
E	36	.135	.529	.114	.886
G	36	.190	.288	.169	.831
S	42	.146	.449	.130	.870
N	36	.083	.709	.089	.911
	150	.139	.492	.126	.874

Table A.10: Variation in experimental results due to metaphor type. Shown are the number of samples N , the effect size (R^2), the p-value statistical significance (** for $p \leq .001$, ** for $p \leq .01$, * for $p \leq .05$, † for $p \leq .1$), the probability of Type I errors α , the probability of Type II errors β , and statistical power $P = 1 - \beta$ ($\diamond = P < .8$). IDs: (1) E mono's and (2) Biling's reading/response times, (3) through (5) are the out-of-context plausibility judgements, (6) through (8) the in-context judgements; (3) E mono, (4) G mono, (5) Biling, (6) E mono, (7) G mono, and (8) Biling. The last row provides general estimates of central tendency, averaged across the individual studies and weighted according to their sample size. (9) Per-participant reading/response/reaction times (RTs), (10) per-item RTs, (11) per-participant out-of-context judgements, (12) per-item out-of-context judgements, (13) per-participant in-context judgements, and (14) per-item in-context judgements.

Per participant					
ID	N	Effect size	$p = \alpha$	β	P
1	13	.394	.0027 **	.687	.313 \diamond
2	12	.445	.0915 †	.667	.333 \diamond
3	13	.683	.000057 ***	.201	.799
4	19	.475	.022 *	.282	.718 \diamond
5	60	.500	.026 *	.001	.999
6	59	.576	.012 *	.001	.999
7	20	.443	.063 †	.305	.695 \diamond
8	21	.529	.098 †	.135	.865
	217	.518	.031 *	.156	.844

Per item					
ID	N	Effect size	$p = \alpha$	β	P
1	25	.394	.0027 **	.250	.750 \diamond
2	30	.445	.0915 †	.083	.917
3	25	.683	.000057 ***	.004	.996
4	18	.475	.022 *	.320	.680 \diamond
5	30	.500	.026 *	.039	.961
6	26	.576	.012 *	.026	.974
7	26	.443	.063 †	.146	.854
8	38	.529	.098 †	.004	.996
	218	.507	.049 *	.094	.906

Weighted average tendencies					
ID	N	Effect size	$p = \alpha$	β	P
9	25	.418	.045 *	.677	.323 \diamond
10	55	.422	.051 †	.159	.841
11	92	.521	.022 *	.087	.913
12	73	.557	.016 *	.096	.904
13	100	.540	.040 *	.090	.910
14	90	.518	.063 †	.051	.949

Table A.11: Variation in experimental results neither explained by metaphor type nor inter-personal variation nor variation between metaphors of the same type. Shown are the number of samples N , the effect size (R^2) of the error term ε , the p-value of statistical significance and probability of Type I errors $\alpha = p(H_1)p(D|H_1) + p(H_2)p(D|H_2) + p(H_3)p(D|H_3)$, the probability of Type II errors β . Statistical power $P = 1 - \beta$ ($\diamond = P < .8$), powers go to 1 as ε goes to 0. IDs: (1) E mono's and (2) Biling's reading/response times (RT), (3) through (5) are the out-of-context plausibility judgements, (6) through (8) the in-context judgements; (3) E mono, (4) G mono, (5) Biling, (6) E mono, (7) G mono, and (8) Biling. The last row provides general estimates of central tendency, averaged across the individual studies and weighted according to their sample size. (9) Per-participant RTs, (10) per-item RTs, (11) per-participant out-of-context judgements, (12) per-item out-of-context judgements, (13) per-participant in-context judgements, and (14) per-item in-context judgements.

Per participant					
ID	N	Effect size	$p = \alpha$	β	P
1	13	.294	.630	.552	.448 \diamond
2	12	.149	.772	.087	.913
3	13	.084	.703	.005	.995
4	19	.215	.640	.068	.932
5	60	.190	.510	.001	.999
6	59	.123	.596	0	1
7	20	.208	.617	.043	.957
8	21	.136	.597	.001	.999
	217	.168	.597	.048	.952

Per item					
ID	N	Effect size	$p = \alpha$	β	P
1	25	.294	.560	.198	.802
2	30	.149	.687	.001	.999
3	25	.084	.651	.001	.999
4	18	.215	.647	.084	.916
5	30	.190	.523	.001	.999
6	26	.123	.576	.001	.999
7	26	.208	.566	.010	.990
8	38	.136	.509	.001	.999
	218	.171	.584	.032	.968

Weighted average tendencies					
ID	N	Effect size	$p = \alpha$	β	P
9	25	.224	.698	.329	.671 \diamond
10	55	.215	.629	.091	.909
11	92	.180	.564	.015	.985
12	73	.160	.597	.021	.979
13	100	.143	.601	.009	.991
14	90	.153	.545	.004	.996