

# **Two of a kind**

Idiomatic expressions in native speakers  
and second language learners

Published by  
LOT  
Kloveniersburgwal 48  
1012 CX Amsterdam  
The Netherlands

phone: +31 20 525 2461

e-mail: [lot@uva.nl](mailto:lot@uva.nl)

<http://www.lotschool.nl>

Cover illustration: Illustrations by Kasper Boon (English Kaboons and Dutch Kaboontjes from 'Het leven is een feest, maar je moet zelf de slingers ophangen', Edition Albert Sickler; text Ingrid Regout and illustrations Kasper Boon). Cover design by Lisa Hubers.

ISBN: 978-94-6093-336-3

NUR: 616

Copyright © 2019: Ferdy Hubers. All rights reserved.

# Two of a kind

*Idiomatic expressions in native speakers and  
second language learners*

Proefschrift

ter verkrijging van de graad van doctor

aan de Radboud Universiteit Nijmegen

op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,

volgens besluit van het college van decanen

in het openbaar te verdedigen op dinsdag 28 januari 2020

om 12.30 uur precies

door

Ferdy Cornelis Wilhelmus Hubers

geboren op 4 maart 1991

te Berlicum

Promotor:

Prof. dr. A.F.J. Dijkstra

Copromotoren:

Dr. H. Strik

Dr. C. Cucchiarini

Manuscriptcommissie:

Prof. dr. R.W.N.M. van Hout

Prof. dr. M. Brysbaert (Universiteit Gent, België)

Prof. dr. K. Conklin (University of Nottingham, Verenigd Koninkrijk)

Prof. dr. A. Housen (Vrije Universiteit Brussel, België)

Prof. dr. A. Weber (Eberhard Karls Universität Tübingen, Duitsland)



The research presented in this dissertation is part of the research program Free Competition in the Humanities with project number 360-70-510 NWO ISLA, which is financed by the Dutch Research Council (NWO).

## Table of Contents

<b>DANKWOORD</b>	<b>xi</b>
<b>CHAPTER 1 GENERAL INTRODUCTION</b>	<b>1</b>
<b>1.1 GOAL OF THIS THESIS AND RESEARCH QUESTIONS</b>	<b>4</b>
<b>1.2 PROPERTIES OF IDIOMATIC EXPRESSIONS</b>	<b>6</b>
<b>1.3 RESEARCH ON IDIOMATIC EXPRESSIONS IN L1 AND L2</b>	<b>8</b>
1.3.1 IDIOM KNOWLEDGE	9
1.3.2 IDIOM LEARNING	13
1.3.3 IDIOM REPRESENTATION AND PROCESSING	15
<b>1.4 THE PRESENT RESEARCH</b>	<b>18</b>
<b>CHAPTER 2 NORMATIVE DATA OF DUTCH IDIOMATIC EXPRESSIONS: SUBJECTIVE JUDGEMENTS YOU CAN BANK ON</b>	<b>21</b>
<b>2.1 INTRODUCTION</b>	<b>24</b>
2.1.1 PROPERTIES OF IDIOMATIC EXPRESSIONS	26
2.1.2 RELIABILITY OF SUBJECTIVE JUDGEMENTS	29
2.1.3 THE PRESENT STUDY	33
<b>2.2 MATERIALS AND METHOD</b>	<b>37</b>
2.2.1 PARTICIPANTS	37
2.2.2 MATERIALS	38
2.2.3 DESIGN AND PROCEDURE	38
2.2.4 DATA ANALYSIS	40
<b>2.3 RESULTS</b>	<b>41</b>
2.3.1 GENERAL RESULTS	41
2.3.2 RELIABILITY	42
2.3.3 IMPACT OF IDIOM PROPERTIES ON OBJECTIVE IDIOM KNOWLEDGE	46
<b>2.4 DISCUSSION</b>	<b>50</b>
2.4.1 RELIABILITY OF SUBJECTIVE IDIOM PROPERTIES	50
2.4.2 COMPARISON BETWEEN SUBJECTIVE AND OBJECTIVE IDIOM FREQUENCY	52
2.4.3 RELATION OF SUBJECTIVE IDIOM PROPERTIES TO OBJECTIVE IDIOM KNOWLEDGE	53
<b>2.5 CONCLUSIONS</b>	<b>55</b>

<b>CHAPTER 3 SECOND LANGUAGE LEARNER INTUITIONS OF IDIOM PROPERTIES: WHAT DO THEY TELL US ABOUT L2 IDIOM KNOWLEDGE AND ACQUISITION?</b>	<b>57</b>
<b>3.1 INTRODUCTION</b>	<b>59</b>
3.1.1 L2 IDIOM ACQUISITION	62
3.1.2 IDIOM PROPERTIES	63
3.1.3 RELIABILITY	68
3.1.4 THE PRESENT STUDY	69
<b>3.2 METHODS</b>	<b>71</b>
3.2.1 PARTICIPANTS	71
3.2.2 MATERIALS	72
3.2.3 DESIGN AND PROCEDURE	73
3.2.4 DATA ANALYSIS	75
<b>3.3 RESULTS</b>	<b>76</b>
3.3.1 RELIABILITY	76
3.3.2 COMPARISON OF L1 AND L2 INTUITIONS	77
3.3.3 INTUITIONS AND OBJECTIVE IDIOM KNOWLEDGE	78
<b>3.4 DISCUSSION</b>	<b>82</b>
3.4.1 RELIABILITY	82
3.4.2 COMPARISON OF L1 AND L2 INTUITIONS	83
3.4.3 INTUITIONS AND OBJECTIVE IDIOM KNOWLEDGE	85
3.4.4 L2 IDIOM ACQUISITION	87
<b>3.5 CONCLUSIONS</b>	<b>88</b>
<b>CHAPTER 4 LEARNING L2 IDIOMS IN A CALL ENVIRONMENT: THE ROLE OF PRACTICE INTENSITY, MODALITY, AND IDIOM PROPERTIES</b>	<b>91</b>
<b>4.1 INTRODUCTION</b>	<b>93</b>
4.1.1 RESEARCH BACKGROUND	95
4.1.2 THE PRESENT STUDY	99
4.1.3 RESEARCH QUESTIONS AND HYPOTHESES	100
<b>4.2 METHOD</b>	<b>101</b>
4.2.1 PARTICIPANTS	101
4.2.2 MATERIAL	102
4.2.3 DESIGN AND PROCEDURE	105

4.2.4 DATA ANALYSIS	107
<b>4.3 RESULTS</b>	<b>107</b>
4.3.1 INTENSITY OF PRACTICE, TRANSPARENCY, AND CROSS-LANGUAGE OVERLAP	107
4.3.2 READING MODALITY	112
<b>4.4 DISCUSSION</b>	<b>115</b>
<b>4.5 CONCLUSIONS</b>	<b>118</b>
<b>CHAPTER 5 IDIOM PROCESSING BY NATIVE SPEAKERS: INDIVIDUAL WORD ACTIVATION AND THE ROLE OF WORD FREQUENCY</b>	<b>119</b>
<b>5.1 INTRODUCTION</b>	<b>121</b>
5.1.1 SEMANTIC EFFECTS OF INDIVIDUAL WORDS DURING IDIOM PROCESSING	122
5.1.2 ORTHOGRAPHIC EFFECTS OF THE INDIVIDUAL WORDS DURING IDIOM PROCESSING	126
5.1.3 THE PRESENT STUDY	128
<b>5.2 EXPERIMENT 1</b>	<b>130</b>
5.2.1 METHODS	130
5.2.2 RESULTS	136
5.2.3 DISCUSSION	138
<b>5.3 EXPERIMENT 2</b>	<b>141</b>
5.3.1 METHODS	141
5.3.2 RESULTS	141
5.3.3 DISCUSSION	149
<b>5.4 GENERAL DISCUSSION</b>	<b>150</b>
<b>5.5 CONCLUSIONS</b>	<b>154</b>
<b>CHAPTER 6 IDIOM PROCESSING BY L2 LEARNERS: THE ROLE OF INDIVIDUAL WORDS AND REPEATED EXPOSURE</b>	<b>157</b>
<b>6.1 INTRODUCTION</b>	<b>159</b>
6.1.1 RESEARCH BACKGROUND	160
6.1.2 CURRENT STUDY	163
<b>6.2 METHODS</b>	<b>165</b>
6.2.1 PARTICIPANTS	165
6.2.2 MATERIALS AND DESIGN	165
6.2.3 PROCEDURE	166

6.2.4 DATA ANALYSIS	166
<b>6.3 RESULTS</b>	<b>167</b>
6.3.1 IDIOM LEARNING	167
6.3.2 IDIOM PROCESSING	170
<b>6.4 DISCUSSION</b>	<b>176</b>
6.4.1 IDIOM LEARNING	176
6.4.2 IDIOM PROCESSING	177
<b>6.5 CONCLUSIONS</b>	<b>180</b>
<b>CHAPTER 7 GENERAL DISCUSSION AND CONCLUSIONS</b>	<b>183</b>
<b>7.1 GENERAL DISCUSSION</b>	<b>184</b>
7.1.1 TO WHAT EXTENT ARE L2 LEARNERS COMPARABLE TO NATIVE SPEAKERS IN TERMS OF KNOWLEDGE, REPRESENTATION, AND PROCESSING OF IDIOMATIC EXPRESSIONS?	184
7.1.2 TO WHAT EXTENT CAN L2 LEARNERS BECOME MORE SIMILAR TO NATIVE SPEAKERS AS A FUNCTION OF PRACTICE?	186
7.1.3 TO WHAT EXTENT ARE L2 LEARNERS SENSITIVE TO THE SAME IDIOM PROPERTIES AS NATIVE SPEAKERS?	187
7.1.4 A COMPREHENSIVE MODEL OF L1 AND L2 IDIOM PROCESSING	189
<b>7.2 SIGNIFICANCE OF THE RESEARCH APPARATUS DEVELOPED: DATABASE, WEB-BASED TEST AND CALL SYSTEM</b>	<b>205</b>
7.2.1 IDIOM DATABASE	206
7.2.2 WEB-BASED TEST OF IDIOM KNOWLEDGE	207
7.2.3 CALL SYSTEM	207
<b>7.3 RELEVANCE FOR SCIENTIFIC RESEARCH</b>	<b>209</b>
<b>7.4 RELEVANCE FOR TEACHING PRACTICE</b>	<b>210</b>
<b>7.5 METHODOLOGICAL RECOMMENDATIONS</b>	<b>211</b>
7.5.1 NORMING STUDIES	211
7.5.2 RESEARCH ON L2 IDIOM PROCESSING	212
<b>7.6 PERSPECTIVES FOR FUTURE RESEARCH</b>	<b>214</b>
<b>7.7 CONCLUSIONS</b>	<b>216</b>
<b>REFERENCES</b>	<b>219</b>
<b>SUPPLEMENTARY MATERIALS</b>	<b>235</b>



<b>NEDERLANDSE SAMENVATTING</b>	<b>253</b>
<b>CURRICULUM VITAE</b>	<b>267</b>
<b>LIST OF PUBLICATIONS</b>	<b>269</b>
<b>LIST OF IDIOMS ON COVER</b>	<b>271</b>



## Dankwoord

Dit proefschrift was nooit tot stand gekomen zonder de hulp en steun, direct en indirect, van een heleboel mensen.

Allereerst wil ik graag mijn (co-)promotoren bedanken. Catia, jouw deur stond altijd open en daar heb ik de afgelopen jaren dan ook dankbaar gebruik van gemaakt. Ik kon voor kleine en grote vragen bij je terecht. Daarnaast had je altijd ideeën voor samenwerking met anderen die het project vooruit konden helpen. Mede dankzij jouw contacten hebben we veel van de data kunnen verzamelen die aan de basis ligt van dit proefschrift. Helmer, bedankt voor de goede discussies die we hebben gehad waarin jij steeds weer een stap terug kon doen om de grote lijn in de gaten te houden. Ook wanneer ik die zelf soms uit het oog verloor. En Ton, ik heb genoten van onze vele discussies over experimenten, de resultaten daarvan en nieuwe mogelijkheden voor onderzoek. Je daagde me steeds uit om verder te denken en het beste uit mezelf te halen.

We hebben aardig wat vergaderd met z'n vieren. Naast dat deze bijeenkomsten vaak productief waren, was het ook nog eens gezellig. Bedankt voor al het vertrouwen dat jullie in mij hebben gehad. Met nog wat publicaties in het vooruitzicht hoop ik nog vaak in deze samenstelling bij elkaar te komen.

Ik wil ook graag even stilstaan bij de andere leden van het ISLA-team, want de waarde van een goede samenwerking en fijne sfeer binnen het team moet niet onderschat worden. Wendy, het was fijn om met jou samen te werken in dit project en om allerlei leuke uitdrukkingen uit te wisselen. Louis, bedankt dat je altijd bereid was om me te op weg te helpen als ik vragen had over statistiek. And Alex, thank you for your discussions about idioms and statistics.

Ook ben ik dank verschuldigd aan iedereen die mij door de jaren heen geholpen heeft met het verzamelen en annoteren van data en het ontwikkelen van het computersysteem. Bas, Elise, Janina, Louisa, Maarten, Marcel, Mesian, Michel, Nino, Romy en Tim, bedankt voor al jullie inspanningen! Steve, thank you for your help and advice with respect to the CALL system you originally developed. Daarnaast wil ik Jan Hoogland bedanken voor het opzoeken van Arabische equivalenten van onze Nederlandse uitdrukkingen.

Een deel van mijn experimenten in dit proefschrift heb ik uitgevoerd in het CLS lab. Ik wil graag Bob en Margret bedanken voor alle ondersteuning.

Dit onderzoek was nooit tot stand gekomen zonder iedereen die heeft deelgenomen aan mijn experimenten. Ik wil Lidy graag bedanken voor de hulp in onze zoektocht naar Duitse T2-leerders van het Nederlands. Ook ben ik Jeanine Coopmans, Manja Mekking, en Tirza van Dongen dankbaar voor de mogelijkheden die ze geboden hebben om de Duitse studenten Logopedie en de Arabische NT2-leerders aan de HAN mee te laten doen met onze leerexperimenten.

Daarnaast ben ik de GSH en het IMPRS dankbaar voor de kansen die ze me geboden hebben om mezelf verder te ontwikkelen gedurende de afgelopen jaren.

Ik wil ook graag mijn paranimfen, Claire en Thijs, bedanken. Claire, gedurende het grootste deel van mijn PhD ben je mijn kantoorgenoot geweest. Ik kijk met veel plezier terug op deze tijd. Ik ben je dankbaar voor de leuke gesprekken die over van alles en nog wat konden gaan en voor de interessante discussies. Er werd natuurlijk hard gewerkt, maar gelukkig was er regelmatig tijd voor een kop koffie. Ik ben nog steeds erg blij met de mok! Thijs, als ik afleiding nodig had, dan kon ik altijd bij jou terecht. Maar ook als ik vragen had, stond je deur altijd open. Je was altijd bereid om mee te denken en teksten te lezen. Bedankt voor alle hulp, de gezellige koffiemomenten, etentjes en de wugs!

Naast Claire heb ik ook nog een lange tijd het kantoor gedeeld met Jan Willem en Saskia. Jan Willem, het was leuk om met je te praten over ons onderzoek. Ook ben ik je dankbaar voor de nodige klustips en het lenen van de afkortaag! Saskia, bedankt voor de gezelligheid en de plantjes op kantoor. Het was fijn om in de laatste fase van mijn PhD een kantoor met jou te delen.

And of course I could not forget my other colleagues from the eighth and ninth floor. You all made me feel at home from the beginning. There was always someone available for questions, complaints, or a chat. I am grateful for all the 'gezellige' coffee breaks, lunch breaks and cake moments. Though these moments would take up much of my time every now and then, I could always go back to work afterwards with 'goede moed' and a 'frisse blik'. Thank you all!

Ook wil ik ook mijn collega's bedanken van de twee onderzoeksgroepen waar ik deel van uitmaakte: Language and Speech, Learning and Therapy, en Cognitive and Developmental Aspects of Multilingualism. Aan de bijeenkomsten van beide groepen heb ik veel gehad. Ik kon altijd rekenen op nuttige feedback en tips.

Daarnaast liep ik ook geregeld binnen bij Roeland als ik een vraag had over statistiek. Roeland, ik waardeer het enorm dat je altijd bereid was om mee te denken en te helpen waar nodig, ook al was je niet mijn begeleider.

Ook wil ik Helen bedanken. Hoewel je niet direct bij mijn proefschrift betrokken bent geweest, ben ik je dankbaar voor de kansen die ik heb gekregen om onderwijs te geven en voor de prettige samenwerking die uiteindelijk heeft geleid tot meerdere artikelen.

Er zijn ook nog een heleboel mensen geweest die niet direct betrokken waren bij mijn onderzoek, maar zonder wie ik het de afgelopen jaren nooit had volgehouden. Cindy en Wouter, bedankt voor jullie interesse, de spelletjesavonden en de lekkere biertjes. Gian, Glen, Joep, Kenny, Koen, Nick, en Ruud, jullie zorgden voor de nodige afleiding in drukke tijden. Bedankt voor de mooie vriendenweekenden en het slap geouwehoer van tijd tot tijd. Mitchell, ik heb menig biertje met jou gedronken en je was altijd benieuwd hoe het met mij en mijn onderzoek ging. Er is veel veranderd in de afgelopen jaren. Waar we eerst met z'n tweeën een biertje gingen drinken bij Café Jos, spreken we tegenwoordig af met z'n zevenen!

Naast vrienden heb ik ook altijd kunnen bouwen op (schoon-) familie. Opa en Oma, jullie vroegen altijd hoe het met mij en mijn onderzoek ging, ook al was het voor jullie niet helemaal duidelijk wat ik nu precies deed. Marcella, Mark, Lyanne, Femke en Eric, bedankt voor de lekkere diners samen, en jullie interesse en gezelligheid. Lisa, lieve zus, bedankt voor het ontwerpen van de kaft van dit proefschrift. Ik heb met jou en Ruud veel gelachen en gezellige avonden gehad. Ik hoop dat er nog veel volgen! Ons pap en mam, bedankt voor jullie onophoudelijke steun, trots en interesse. Ik weet dat ik altijd op jullie kan rekenen. Bedankt.

En natuurlijk lieve Renske, mijn rots in de branding. Bedankt dat je altijd voor me klaar staat en bereid bent me te helpen met wat dan ook. Bedankt ook voor je geduld tijdens de afronding van dit proefschrift. De

afgelopen tijd hebben we samen veel mooie momenten beleefd met als kers op de taart de geboorte van Teije. Ik kijk uit naar nog veel meer mooie en bijzondere herinneringen samen.

En tot slot, Teije. Je bent er nog maar net, maar bedankt dat je mijn stok achter de deur was om dit proefschrift uiteindelijk echt op tijd af te hebben. Ik kan me geen leven meer voorstellen zonder jou.

## Chapter 1

### **General Introduction**

## 2 | Chapter 1

Everyday language is characterized by recurrent sequences of words that tend to co-occur in more or less fixed combinations. Examples are phrases like *hit the sack*, *make a decision*, *plastic surgery*, *back and forth*, *in the middle of*, *I beg your pardon*. The special characteristic of such phrases is that they consist of fixed combinations of words that are not produced on the fly, but have become entrenched and have assumed the character of formulas, hence the term formulaic expressions (FEs).

Formulaic expressions come in different types. Categories of FEs that have been distinguished in the literature are idiomatic expressions (*hit the sack* and *be two of a kind*), collocations (*plastic surgery*, *heavy rain* and *make a decision*), binomials (*back and forth*, *black and white*, and *man and wife*), lexical bundles or polywords (*by the way*, *to begin with*, *in any case*), phrasal verbs (*to give up* and *to put up with*), and speech formulas (*you're welcome* and *I beg your pardon*) (Cacciari, 2014; Nattinger & DeCarrico, 1992; Siyanova-Chanturia & Martinez, 2014; Sprenger, Levelt, & Kempen, 2006; Wray, 2008).

Examples of collocations are *make a decision* and *plastic surgery*. Both phrases are fixed sequences of words. In order to express the meaning of *make a decision*, the verb *make* should be used in combination with *a decision*, because this sequence is more frequently used by native speakers rather than its atypical counterpart *take a decision*. Due to the fixed character of *make a decision*, any deviation from this sequence is considered to be atypical. The same goes for *plastic surgery*. This word string frequently occurs in English, whereas the combination *plastic operation* does not. The latter sequence is considered to be atypical, because it is hardly used by native speakers, even though it is perfectly grammatical and conveys essentially the same meaning as its typical counterpart.

For binomials not only the combination of words is important, but also their order. Reversing the order of adverbs in the reversible binomial *back and forth*, for example, results in the atypical combination *forth and back*. Again, the meaning is the same as that of its typical counterpart and the sequence is perfectly grammatical. However, *forth and back* is hardly used by native speakers of English and therefore is considered to be odd in English. In addition to reversible binomials, also irreversible binomials exist, such as *by and large*. For this type of binomials, the word order and



the meaning are so conventionalized that reversing the words leads to an anomalous meaning. The number of irreversible binomials is relatively small.

Speech formulas tend to be used in specific situations only and/or have special communicative functions (Aijmer, 1996; Erman, 2007; Kecskés, 2000). The phrase *I beg your pardon*, for example, is a combination of words that is typically used as an indirect request for clarification or to apologise. The expression *can you clarify that* can also be used as a request for clarification, but in specific situations native speakers more often opt for *I beg your pardon*. The meaning of this chunk is similar to the meaning of the direct request for clarification.

Replacing words in the examples above by synonyms results in atypical speech that is hardly used by native speakers, while the meaning of the phrases does not change. Replacing the word *strong* with its synonym *powerful* in the string *strong coffee*, for example, does not necessarily change its meaning, but leads to a word combination that native speakers of English would not use. A category of formulaic expressions for which this operation does lead to changes in meaning is idiomatic expressions. Idiomatic expressions are usually defined as recurrent sequences of words that convey a figurative meaning (Abel, 2003; Cacciari & Glucksberg, 1991; Titone, Columbus, Whitford, Mercier, & Libben, 2015). An example is *hit the sack*. This phrase consists of a fixed sequence of words, and can have two meanings. Depending on the context, it can be used in a literal or in a figurative sense. The figurative meaning of *hit the sack* ('go to bed') can only be expressed by this exact combination of words. In this case it is not possible to use the synonym *bag* instead of *sack*, because the resulting phrase *hit the bag* can only convey the literal meaning. While all idiomatic expressions have a figurative meaning, not all idiomatic expressions necessarily have a literal meaning. An example of an idiom that only can be interpreted figuratively is *shoot the breeze* ('have an informal conversation'): One cannot literally shoot a breeze.

Formulaic expressions are pervasive in language (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Erman & Warren, 2000; Pollio, Barlow, Fine, & Pollio, 1977; Sorhus, 1977). Sorhus (1977) analyzed a corpus of spoken Canadian English and calculated that people

use one formulaic expression every five words, and according to Erman and Warren (2000), about 52% of written and 58% of spoken language is formulaic. These studies indicate that formulaic expressions are frequently used by native speakers. For idiomatic expressions, the general finding is that as a group they occur frequently in daily language (Ellis, 2012), whereas individual idioms are not so frequent (Ellis & Wulff, 2015).

Because idiomatic expressions constitute a substantial part of the daily language of native speakers, it is important that second language (L2) learners acquire this type of language. Research has shown that L2 learners come across as more proficient if they use idiomatic expressions and other types of formulaic language (Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006) and L2 learners consider learning L2 idioms important (Liontas, 2015b). However, L2 learners appear to use fewer idiomatic expressions than native speakers (Kecskes, 2007, 2015). Kecskes (2007), for example, analyzed conversations of non-native speakers of English. He found that only 7.6% of their language was formulaic and that only 0.8% of their language contained idiomatic expressions. Differences between native speakers and L2 learners have also been observed in the processing of idioms. Research has shown that L2 learners who acquire their L2 late in life comprehend and produce idiomatic expressions differently than native speakers (Titone et al., 2015). Although idiomatic expressions are pervasive in native language, L2 learners are found to experience enormous difficulties in mastering this type of language (Cieślicka, 2006; Conklin & Schmitt, 2008; Ellis, Simpson-Vlach, & Maynard, 2008; Kovecses & Szabó, 1996; Wray, 2000). Even highly proficient L2 learners experience difficulties understanding and using idiomatic expressions (Ellis et al., 2008).

### **1.1 Goal of this thesis and research questions**

The variation and complexity of idiomatic expressions can be confusing for L2 learners (Pawley & Syder, 1983). At present, it is not clear how L2 learners deal with idiomatic expressions in their new language and how their performance and processing compare to those of native speakers. How do these learners process idiomatic expressions on-line, and what

are the mental representations underlying them? And how does repeated exposure to and focused practice with L2 idioms affect L2 idiom acquisition and processing? These are important issues that in the end co-determine how proficient and native-like a learner is in the new language. While these questions are clearly interconnected, previous research has tended to address them separately. The research reported on in this thesis attempts to address these questions in an integrated manner by adopting a multidisciplinary approach to idiomatic expressions in both the L1 and the L2. The aim of this research is to investigate the knowledge, representation, learning and processing of idiomatic expressions by second language learners relative to native speakers. More specifically, this thesis addresses the following general research questions.

First, to what extent do L2 learners compare to native speakers in terms of idiom knowledge, representation, and processing? Answers to this question may provide insights into underlying mechanisms of L2 learners and native speakers during idiom comprehension and production. Do L2 learners approach idiomatic expressions in a fundamentally different way from native speakers or are similar mechanisms at work?

Second, to what extent can L2 learners become similar to native speakers in terms of idiom knowledge, representation, and processing as a function of learning? More specifically, is there a role for intensive practice with and repeated exposure to L2 idiomatic expressions? Can differences observed between L2 learners and native speakers be overcome by sufficient exposure to the L2 in general and L2 idioms in particular?

Finally, to what extent do the properties of idioms affect idiom knowledge, representation, learning and processing in L2 learners relative to native speakers?

In order to address the first question, we examine intuitions about idiom properties by native speakers and objectively assess idiom knowledge in native speakers (Chapter 2). In addition, we compare idiom property intuitions and idiom knowledge in native speakers to those in L2 learners (Chapter 3), and investigate L1 and L2 idiom processing (Chapters 5 and 6). To answer the second research question, we

investigate changes in idiom knowledge (Chapter 4), and idiom processing (Chapter 6) as a function of focused practice through the use of a Computer Assisted Language Learning (CALL) system. The third question is investigated in all Chapters. In the Chapters 2, 3 and 4, we examine to what extent idiom properties are related to L1 and L2 idiom knowledge, while in Chapters 5 and 6 we look into the effects of idiom properties on idiom processing.

Before presenting these studies in more detail, we first discuss the idiom properties that are investigated in this thesis (Section 1.2). In Section 1.3, we provide an overview of the research that has investigated idiom knowledge, learning, processing, and representation in both the L1 and L2. Finally, we conclude this Chapter by presenting an outline of the thesis in Section 1.4.

## 1.2 Properties of idiomatic expressions

Idiomatic expressions can vary along different dimensions (Cieślicka, 2015; Titone et al., 2015) that have been shown to affect idiom knowledge, learning and processing. Dimensions can, for example, be related to the experience a speaker has with the idioms under consideration, i.e., *frequency* and *familiarity*. Some idioms are less frequent in daily speech than others and therefore they are less familiar to native speakers. The idiom *out of the blue* is considered to be highly familiar and is quite frequently used. Therefore, this idiomatic expression will be easier to understand than unfamiliar and infrequent ones, such as *a pig in a poke* (Titone et al., 2015).

Other dimensions that are distinguished in the literature on idiomatic expressions are *literal plausibility*, *semantic decomposability*, and *transparency* (Cieślicka, 2015; Steinel, Hulstijn, & Steinel, 2007; Titone & Libben, 2014). Literal plausibility is defined as the possibility for an idiom to have a literal interpretation. For some idioms the literal interpretation can be extracted quite easily (e.g. *kick the bucket*, and *spill the beans*), whereas other idiomatic expressions cannot easily be interpreted in a literal way (e.g. *shoot the breeze*) (Titone & Libben, 2014). Semantic decomposability is concerned with the contribution of the individual word meanings to the overall figurative meaning (Cieślicka,

2015). The expression *spill the beans* is semantically decomposable, because the individual word meanings map onto different parts of the figurative meaning: *spill* corresponds with *divulge something* and *the beans* map onto *secret*. For *kick the bucket*, no such direct mapping exists; thus, this idiomatic expression is semantically non-decomposable. Transparency is related to semantic decomposability, but transparency refers to the extent to which the figurative interpretation can be extracted from the literal one (Cieślicka, 2015). The expression *spill the beans*, which is semantically decomposable, is rather non-transparent or opaque, because the figurative meaning *to reveal a secret* cannot be extracted from the literal interpretation. The idiomatic expressions *to hit two birds with one stone*, on the other hand, can be seen as transparent, because the figurative meaning (*to solve two problems at once by a single action*) can be extracted on the basis of the literal interpretation. Although the terms *semantic decomposability* and *transparency* have sometimes been used interchangeably to refer to the same property (Cieślicka, 2015), it is important to realize that they are two distinct notions.

Another dimension along which idiomatic expressions can vary is *imageability* (Steinel et al., 2007), often defined as the extent to which an idiom can be visualized (Cacciari & Glucksberg, 1995; Steinel et al., 2007). The English idiom *break the ice* can evoke a mental image of someone breaking the ice and therefore is highly imageable. An idiom such as *to hang fire* is hardly imageable. It is important to note that either the literal or the figurative meaning can be visualized. Cacciari and Glucksberg (1995), for example, found that native speakers tend to form a mental image of the idiom's literal meaning rather than its figurative meaning.

*Cross-language overlap*, or *cross-language similarity* is a property of the idiom that is relevant for L2 learners only, and refers to the extent of overlap in form and meaning between idioms in the L1 and the L2 of a language learner. Some idiomatic expressions in one language are identical in form and meaning to idioms in the other language. The English idiom *to play the first fiddle* also exists in Dutch as a direct translation: *de eerste viool spelen* (lit. 'to play the first fiddle'). However, expressions can also be different between two languages. The German equivalent of the English idiom *to bite the dust* is *ins Gras beißen* (lit. 'to

bite in the grass’). Another example comes from English and Dutch. The Dutch equivalent of the English idiom *to grab the bull by the horns* is *de koe bij de hoorns vatten* (lit. ‘to grab the cow by the horns’). With respect to the idiomatic meaning these expressions are identical, but they are slightly different in form. It is also possible that the idiomatic meaning in two languages is identical, but that the corresponding idiomatic expressions are completely different with respect to their form. The Dutch version of the German idiom *nicht alle Tassen im Schrank haben* (lit. ‘to not have all the cups in cupboard’), for example, is *een steekje los hebben* (lit. ‘to have a stitch loose’). Both idioms have the same meaning (‘to be crazy’). However, this is expressed in totally different words. The expressions described above are all identical in terms of their meanings. However, an idiom can also exist in one language, but not in the other. The Dutch idiomatic expression *goed uit de verf komen* (lit. ‘to come well out of the paint’, meaning ‘to stand out well’) does not exist as an expression in English. If the L1 and the L2 of an L2 learner are relatively closely related, the languages are likely to have a high degree of cross-language overlap, while this is likely to be less if the L1 and the L2 of an L2 learner are more distant.

### **1.3 Research on idiomatic expressions in L1 and L2**

Idiomatic expressions have been topic of investigation in different sub-disciplines within linguistics, such as applied linguistics and psycholinguistics.

In applied linguistics, idiomatic expressions are often seen as units that are comparable to single words and have a special function in discourse (Boers & Lindstromberg, 2012). Research on idiomatic expressions within this field has mainly focused on the level of idiom knowledge and use by L2 learners, the difficulties they experience with L2 idiom comprehension and production (Irujo, 1993; Kellerman, 1979; Pawley & Syder, 1983), the factors that influence L2 idiom comprehension, production and learning (Abel, 2003; Cooper, 1999; Steinel et al., 2007) and the effectiveness of teaching and learning methods to improve the comprehension and production of L2 formulaic expressions in general and L2 idioms in particular (Boers, Eyckmans,

Kappel, et al., 2006; Boers, Eyckmans, & Stengers, 2007; Boers & Lindstromberg, 2012; Stengers, Deconinck, Boers, & Eyckmans, 2016; Türker, 2016; Zyzik, 2011).

Due to their (semi) fixed character, idioms are often treated as single units. In order to find out whether this indeed is the case, psycholinguistics has studied the processing of idioms and more specifically, whether these expressions are stored and retrieved as a whole from the mental lexicon. This issue has been extensively studied in native speakers (Bobrow & Bell, 1973; Cacciari & Tabossi, 1988; Cutting & Bock, 1997; Rommers, Dijkstra, & Bastiaansen, 2013; Sprenger et al., 2006; Swinney & Cutler, 1979), and more recently also in L2 learners (Beck & Weber, 2016a; Carrol & Conklin, 2014; Cieślicka, 2006, 2013; van Ginkel & Dijkstra, 2019). In addition, researchers have investigated to what extent the idiom properties discussed above affect idiom processing (Libben & Titone, 2008; Titone et al., 2015; Titone & Libben, 2014).

Idiomatic expressions have received much attention in the literature and have been studied from many different perspectives producing mixed results. The heterogeneity of idiomatic expressions might be one of the reasons for these mixed results. The next subsections give a brief overview of research that studied the knowledge (Section 1.3.1), learning (Section 1.3.2) and processing (Section 1.3.3) of idiomatic expressions in native speakers and L2 learners.

### **1.3.1 Idiom knowledge**

Idiom knowledge has been studied in both native speakers and L2 learners. Data on native speaker idiom knowledge have been mostly collected as part of large scale norming studies, in which also subjective judgements have been obtained for various properties of idiomatic expressions (Bonin, Méot, & Bugajska, 2013; Citron et al., 2016; Libben & Titone, 2008; Tabossi, Arduino, & Fanari, 2011). This information is often used to select appropriate idiomatic expressions for psycholinguistic experiments and to investigate how idiom properties affect L1 and L2 idiom processing and idiom learning. In some cases researchers also examined to what extent idiom properties, such as frequency and transparency affect idiom knowledge.

Idiom knowledge has been assessed in different ways in the literature. The most commonly used operationalization is by asking participants to what extent they are familiar with (the meaning of) the idiomatic expression as rated on a 5- or 7-point scale (Bonin et al., 2013; Libben & Titone, 2008).

Libben and Titone (2008), for example, investigated the impact of idiom properties on meaningfulness judgements of English idiomatic expressions. They operationalized meaningfulness as the extent to which participants considered the phrase to be meaningful, and took this to represent the participants' "confidence in their understanding of what the phrase actually means" (Libben & Titone, 2008, p. 1106). This can be seen as an indirect and subjective measure of receptive idiom knowledge. In general, native speakers appeared to be familiar with the meaning of idiomatic expressions. Moreover, frequency positively affected idiom knowledge, and semantic decomposability had a positive effect on idiom knowledge only for infrequent idioms.

Only a few studies have measured idiom knowledge in a more objective way (e.g., Carrol, Littlemore, & Gillon Dowens, 2017; Tabossi et al., 2011). These studies mainly tested receptive idiom knowledge. Tabossi et al. (2011), for example, assessed receptive idiom knowledge by asking native speakers to provide the meaning of 245 Italian idioms in an open question. In addition, the authors collected data on various idiom properties, such as literality and transparency, and investigated to what extent idiom properties affected L1 idiom knowledge. They showed that, overall, idiom knowledge was rather high (80% correct), but that much variation existed among native speakers. Subjective familiarity and transparency turned out to positively affect idiom knowledge.

The studies above suggest that the relative transparency, frequency, and familiarity of idiomatic expressions are important predictors of L1 idiom knowledge.

The first studies that looked into L2 idiom knowledge mainly focused on the role of L1 knowledge (Irujo, 1986b, 1993; Kellerman, 1979). Irujo (1986) tested receptive and productive knowledge of idiomatic expressions by advanced Venezuelan learners of English, who were native speakers of Spanish. She identified three types of idiomatic expressions: expressions that were identical in form and meaning in the



two languages (identical), expressions that were identical in meaning, but only similar in form (similar), and expressions that were identical in meaning, but completely different in form (different). Receptive idiom knowledge was tested in comprehension tasks by asking the participants to select the idiom's meaning in a multiple-choice exercise (meaning recognition) and to describe it in an open question (meaning recall). To assess productive idiom knowledge, participants had to perform a discourse completion task and a translation task (production tasks). Irujo (1986) found that the participants' native language affected both receptive and productive L2 idiom knowledge. In the comprehension tasks (receptive idiom knowledge), participants performed significantly better in the case of identical and similar idioms as compared to different idioms. The two production tasks (productive idiom knowledge) revealed that it was much easier to produce identical idioms than the other types of idioms. L1 interference during production was found more often for similar idioms than for different idioms. This study suggests that cross-language overlap may facilitate comprehension of L2 idioms that are identical in form and meaning and idioms that are slightly different in form, while idiom production may be facilitated only for idioms that are completely identical in form and meaning in the two languages. The same pattern of results was found for more proficient Spanish speakers of English (Irujo, 1993) and for Malay speaking learners of English (Charteris-Black, 2002).

In addition to the role of cross-language overlap, the effect of other idiom properties on L2 idiom knowledge has been investigated (e.g., Abel, 2003; Carrol et al., 2017; Cooper, 1999). Cooper (1999) investigated what strategies L2 learners employ during the comprehension of idiomatic expressions. The participants were asked to orally give the meanings of 20 idiomatic expressions (receptive idiom knowledge) and to verbalize their thoughts during the task (think-aloud procedure). The idioms were incorporated in a context consisting of one or two sentences. If participants were unable to provide the correct meaning of an idiom, this was mainly because one of the component words was not known to them, or because they were unable to extract the figurative meaning from the literal meaning due to its opaque character. The idiomatic expressions that were considered to be the easiest were

expressions that the participants indicated to be highly frequent in daily life (frequency of exposure). The strategies that were mostly used to identify the meaning of the idiomatic expressions were guessing on the basis of the context, analyzing the individual words, and using the literal meaning of the idiom.

Abel (2003) focused on the effect of semantic decomposability and frequency on subjective idiom knowledge (familiarity). She conducted a decomposability judgement task in which L2 learners of English had to identify whether the idiomatic expression was semantically decomposable or not. Subsequently the participants were asked to rate the familiarity of the idioms on a 7-point scale. The results were compared to the results of a similar study carried out with native speakers of English (Titone & Connine, 1994). The results of the decomposability judgement task were found to be different for native speakers and L2 learners of English. L2 learners judged more idioms to be decomposable as compared to the native speakers, even if the idiomatic expressions were non-decomposable. On the basis of these results, Abel (2003) suggested that L2 learners tend to rely more on the literal meanings of the individual words when accessing the meaning of L2 idioms. In order to theoretically account for these data, Abel developed a model of L2 idiom comprehension (Abel, 2003). This Dual Idiom Representation Model specifically focused on the role of familiarity and semantic decomposability. According to this model, the way an idiomatic expression is represented in the mental lexicon depends on its decomposability. Abel (2003) suggested that non-decomposable idiomatic expressions have a separate lexical entry, while decomposable idiomatic expressions are considered to be computed on the basis of their individual constituents and therefore do not have a separate lexical entry in the mental lexicon of L2 learners. However, L2 idioms that are encountered more frequently can develop their own entry over time. Within this model, the frequency or familiarity of L2 idiomatic expressions is taken to account for the differences between native speakers and L2 learners: Native speakers have developed more idiom entries because they have been more frequently exposed to these expressions than L2 learners.

In a rating study, Carrol, Littlemore, and Gillon Dowens (2017) collected information on idiom knowledge from native speakers and L2 learners of English. Receptive idiom knowledge was assessed in a multiple-choice question in which participants had to select the correct meaning out of four alternatives. They found that the English idioms were well-known by native speakers (88% correct), while L2 learners performed worse (60% correct). Subjective familiarity and transparency turned out to positively affect idiom knowledge in both native speakers and L2 learners.

The majority of studies investigating idiom knowledge of L2 learners have assessed receptive idiom knowledge. Overall, these studies show that the L2 idiom knowledge is sensitive to the same idiom properties as that of native speakers. In addition, cross-language overlap is found to affect L2 idiom knowledge.

### **1.3.2 Idiom learning**

Early research on the comprehension and production of idiomatic expressions by L2 learners already acknowledged the difficulties L2 learners experience in mastering idiomatic language (Pawley & Syder, 1983). Factors that could possibly influence the learning of idiomatic expressions have been investigated extensively over the years (Boers et al., 2007; Boers, Píriz, Stengers, & Eyckmans, 2009; Steinel et al., 2007; Stengers et al., 2016; Türker, 2016; Zyzik, 2011).

Steinel and colleagues (2007) assessed the effect of transparency, imageability, and direction of learning and testing on L2 idiom learning. The authors conducted a paired-associate learning (PAL) task in which Dutch university students were taught twenty English idiomatic expressions. The PAL paradigm consisted of a training and a test phase. During the training phase the participants were presented with either the L1 idiom first and subsequently its L2 translation (productive learning; L1-L2) or the other way around (receptive learning; L2-L1). In the test phase participants were shown either the L1 idiom (productive testing; L1-L2) or the L2 idiom (receptive testing; L2-L1) and were asked to type in its translation. The L2 idioms varied in transparency and imageability. The authors found that direction of learning affected performance. Participants who had learned the idioms productively (L1-

L2) performed better on the productive test (L1-L2) than participants who had learned the idioms receptively (L2-L1). In addition, productive learners did not outperform receptive learners on the receptive test. Imageability was found to increase performance in receptive learners, but to a lesser extent in productive learners. Transparency was found to have a facilitative effect on performance in the receptive test: High transparent idioms were comprehended better than low or intermediate ones, whereas this effect of transparency was not observed in the productive test. The same facilitative effect of transparency was observed in Skoufaki (2008) for unfamiliar L2 idiomatic expressions. She found that it was easier for L2 learners to comprehend transparent unfamiliar idiomatic expressions than non-transparent unfamiliar ones.

Other researchers have investigated the pedagogical aspect of L2 idiom learning (Boers et al., 2007; Stengers et al., 2016; Türker, 2016; Zyzik, 2011). In a series of experiments, Boers and colleagues investigated the pedagogy of idiom learning and teaching using a web-based application (Boers, Demecheleer, & Eyckmans, 2004; Boers et al., 2007, 2009). Boers et al. (2007), for example, investigated whether knowledge of the etymology of L2 idiomatic expressions aided learning. The experiment was carried out with Dutch learners of English. During a test participants were asked to answer five questions per idiomatic expression. A control group of participants had to identify the meaning of the idiom and subsequently the source, whereas the experimental group was asked to identify the source first and subsequently the meaning. After identifying the source, participants received feedback consisting of the right answer and an elaboration on the etymology of the idiom. Boers and colleagues (2007) observed that participants who first had to identify the source more accurately identified the meaning of the idiom than participants who had not received the source information first.

Türker (2016) investigated to what extent cross-language overlap and supportive context affect idiom acquisition. English L2 learners of Korean participated in an experiment consisting of a pre-test, a Computer Assisted Language Learning (CALL) session, and a post-test. The idioms included in the experiment were either (1) identical in form and meaning in the L1 and L2; (2) identical in form, but different in meaning in the L1 and L2; or (3) they existed in the L2 only. During the

learning session the idioms were presented in different elaborate contexts and participants had to complete various types of exercises. Türker (2016) found that the effect of cross-language overlap was mitigated when supportive context was present in the instructional input. More specifically, if the input included enough context, idioms that existed only in the L2 were learned as well as idioms that also existed in the L1.

### 1.3.3 Idiom representation and processing

Researchers that have investigated idiom processing and representation were mainly interested in the processes underlying idiom comprehension and production and the status of idiomatic expressions in the mental lexicon. More specifically, the question addressed was whether idiomatic expressions have a separate entry in the mental lexicon or whether the figurative meaning is computed from the idiom's component words. This question has been extensively studied in native speakers (Bobrow & Bell, 1973; Cacciari, 2014; Cacciari & Tabossi, 1988; Libben & Titone, 2008; Rommers et al., 2013; Swinney & Cutler, 1979; Titone & Libben, 2014), while research started to address idiom processing in L2 learners only recently (Beck & Weber, 2016a; Carrol & Conklin, 2014, 2017; Cieślicka, 2006; Titone et al., 2015; van Ginkel & Dijkstra, 2019; Yeganehjo & Ngee Thai, 2009).

In the native idiom processing literature roughly three different models of idiom comprehension have been distinguished: (1) non-compositional models, (2) compositional models, and (3) hybrid models. Non-compositional models of idiom processing propose that idioms are stored as a whole in the mental lexicon and that the individual words are not activated during processing (Bobrow & Bell, 1973; Gibbs, 1980). One of the first non-compositional models of idiom processing was developed by Bobrow and Bell (1973). According to their Idiom List Hypothesis, idiomatic expressions were stored in a special list that is not part of the mental lexicon. This idiom list could be accessed via a special *idiom mode* of processing. The idea was that the expressions would undergo a literal analysis first and subsequently be processed in the idiom mode, which predicted a processing advantage for the literal analysis over the figurative analysis.

However, later studies showed that individual words in idiomatic expressions did contribute to their figurative interpretation, leading to compositional models of idiom processing (Cacciari & Glucksberg, 1991; Cacciari & Tabossi, 1988; Gibbs & Nayak, 1989; Nunberg, 1979). Cacciari and Tabossi (1988) found that participants were faster in response to the literal meanings of non-predictable idioms (no bias towards the idiomatic completion of a string), than to their idiomatic meaning. For predictable idioms, the idiomatic meaning was more quickly recognized. The first pattern of results (literal meaning was recognized faster than idiomatic meaning) could be explained by the Idiom List Hypothesis, but the latter pattern of results (idiomatic meaning was recognized faster than literal meaning) could not. In order to explain both patterns of results, Cacciari and Tabossi (1988) proposed the Configuration Hypothesis. This hypothesis did not consider idioms to have separate entries in the mental lexicon. Instead, the same lexical items would be activated during the processing of the idiomatic meaning as during the processing of the literal meaning. This meant that the literal meaning would become available first. The idiomatic meaning would be available after information that identified the phrase as an idiom (the idiom's key) had been processed. The interpretation of the idiom would be literal until the configuration has been recognized. In the case of predictable idioms or an idiomatic biasing context, the configuration would be recognized earlier, which led participants to respond faster to the idiomatic than the literal meaning, whereas in the case of non-predictable idioms or literal biasing contexts, the configuration would be recognized later, which led to the exact opposite pattern.

Finally, hybrid models of idiom processing suggest that idiomatic expressions can be processed compositionally as well as non-compositionally. The literal and idiomatic processing run in parallel and depending on the moment in time and the idiom properties, individual words are activated or not (Beck & Weber, 2016a; Cutting & Bock, 1997; Libben & Titone, 2008; Sprenger et al., 2006; Titone et al., 2015; Titone & Libben, 2014). For example, Sprenger and colleagues (2006) proposed a hybrid model of idiom production. In this model, idiomatic expressions have separate representations (superlemmas). The superlemma is connected to the simple word lemmas that the superlemma is made up

of. These connections express an *element-of* relationship between the simple lemmas and the superlemmas, whereas the connections between the lemmas and the concepts express a meaning relationship. The superlemma that represents the idiom *hit the road*, for example, is connected to the single lemmas *hit*, *the* and *road* on the lemma level. This *element-of* relationship allows for co-activation of the simple lemmas if the superlemma is activated. The lemmas and superlemmas are in turn connected to their corresponding concepts via a meaning relationship, which does not allow co-activation. This means that activating the concept of *hit the road* as an idiom does not activate the individual lemmas directly. By adopting the notion of superlemmas, idiom retrieval becomes similar to the retrieval of single words. According to another hybrid model, the Constraint-Based Model proposed by Libben and Titone (2008), idiom processing is affected by different kinds of information at different points in time. Factors such as familiarity and predictability, which are related to direct retrieval, may affect early stages of idiom comprehension, whereas decomposability or transparency may become important later (Titone et al., 2015; Titone & Libben, 2014).

Although L2 idiom processing has been studied less extensively than L1 idiom processing, various models have been developed to account for L2 idiom processing. One of these models is Cieśllicka's Literal Salience Model (Cieśllicka, 2006). Cieśllicka adopted the notion of salience from Giora (1997), who defined salient meanings as the meanings that are activated first and most strongly during processing, regardless of contextual bias. L2 learners who acquire their L2 in a formal setting tend to know the literal meaning of words before they encounter these words as part of an idiomatic expression (Abel, 2003; Cieśllicka, 2006, 2015). The Literal Salience Model, therefore, considered the literal meaning of idiomatic expressions to be more salient than their figurative meaning. This was expected to be the case for both semantically decomposable and semantically non-decomposable idiomatic expressions. In order to test her model, Cieśllicka (2006) employed a cross-modal priming experiment in which advanced Polish learners of English were auditorily presented with an English sentence containing an idiom. The idioms had all been rated to be familiar by a comparable group of L2 learners. During the

presentation of the sentence, a target word appeared on the screen and participants had to decide whether this word was an existing word of English or not. It was found that the advanced L2 learners were faster in response to a target word that was related to the literal meaning of the idiom than in response to a target word that was related to the figurative meaning. Cieślicka took this processing advantage as evidence in favor of her Literal Salience Model.

Titone et al. (2015) investigated the effect of familiarity and decomposability, which are found to facilitate L1 idiom processing (Libben & Titone, 2008), and the effect of cross-language overlap on the processing of L2 idiomatic expressions. The researchers presented French learners of English with sentences containing idiomatic expressions and literal sentences. The idiom-final word could be presented in English (neutral condition) or in French (code-switching condition). Participants were asked to decide whether these sentences were meaningful or not. Titone and colleagues (2015) found longer decision times for sentences containing idiomatic expressions than for literal sentences. Idiomatic sentences in the code-switching condition were reacted to more slowly than literal sentences in this condition. In addition, highly familiar idioms were reacted to faster than unfamiliar idioms in both the sentences ending in an English and a French noun. A similar facilitatory effect was found for idioms with a high degree of cross-language overlap as opposed to idioms with a low degree of cross-language overlap. However, this effect only reached significance in the code-switching condition. These results led Titone et al. (2015) to conclude that familiarity may facilitate L2 idiom comprehension, which is consistent with models of L1 idiom comprehension (Libben & Titone, 2008). In addition, cross-language overlap was also found to facilitate L2 idiom processing. Semantic decomposability, which was observed to facilitate idiom processing in native speakers, did not seem to play a role in L2 idiom processing.

#### **1.4 The present research**

The research reported on in this thesis aims to increase our understanding of how L2 idioms are known, represented, learned, and



processed by L2 learners as compared to native speakers. In doing so it adopts a multidisciplinary approach that combines methods from applied linguistics and psycholinguistics. In order to gain a deeper understanding of how L2 learners deal with idiomatic expressions in their second language, it is important to know how native speakers understand and process this type of language in the first place. To this end, we first investigate the receptive knowledge of idiomatic expressions and the intuitions about idiom properties in native speakers of Dutch in a large-scale rating study (Chapter 2). The aim of the study presented in **Chapter 2** is to provide information about the extent to which Dutch idiomatic expressions are known and used by Dutch native speakers, whether their intuitions about important properties of idioms can be trusted, and to what extent idiom properties affect idiom knowledge in native speakers. As a result, the data obtained in this study constitute an important basis for selecting materials and for defining benchmarks for the research reported on in the other chapters.

After having established benchmarks for the knowledge of idiomatic expression by native speakers, we go on to investigate the same constructs in intermediate to highly proficient German L2 learners of Dutch using a similar paradigm (Chapter 3). In **Chapter 3**, we examine to what extent L2 learners are able to develop reliable intuitions about properties of L2 idioms, and we compare their intuitions and receptive idiom knowledge to those of native speakers. The choice for German learners of Dutch throughout this thesis was motivated by a number of reasons. First, German and Dutch are very closely related languages, so they can be seen as one of the two extreme points on the continuum of cross-language similarity, which constitutes a good starting point for studying the effects of cross-language similarity on idiom processing. A second important reason is the vicinity of Germany to the Netherlands and the close relationships between the two countries which should stimulate interchange and, possibly, more natural language acquisition, which is known to be particularly beneficial for acquiring formulaic language. A third concomitant reason is the relatively large presence of German students in Nijmegen, a Dutch city close to the German border, and the consequent ease of finding participants for our experiments.

The study on knowledge and intuitions of German L2 learners about Dutch idioms allows us to examine to what extent differences in experience with Dutch between native speakers and L2 learners lead to different intuitions about idiom properties and to differences in idiom knowledge.

The question then is to what extent potential differences in idiom knowledge between native speakers and L2 learners can be resolved by focused training. This question is addressed in **Chapter 4**. It describes a learning study in which German L2 learners of Dutch practiced with L2 idiomatic expressions using a Computer Assisted Language Learning (CALL) system providing immediate feedback. By manipulating the intensity of practice it is possible to examine to what extent repeated exposure can be effective. In addition, by choosing the learning materials properly, the effects of idiom properties on idiom learning can also be assessed.

Once L2 learners have been intensively exposed to L2 idioms, it is possible to make comparisons between native speakers and L2 learners in terms of processing. A first study reported on in **Chapter 5** investigates the role of the individual words during the processing of opaque idiomatic expressions in native speakers. A second study presented in **Chapter 6** combined approaches from applied linguistics and psycholinguistics to investigate the same phenomenon in L2 learners. German L2 learners of Dutch that had received focused training through the CALL system (Chapter 4), participated in the same experiment as conducted with native speakers (Chapter 5). This allows us to investigate the effect of individual word activation and the effect of repeated exposure on the processing of opaque idiomatic expressions by L2 learners.

In **Chapter 7**, we present the results of the research reported on in this thesis from a general perspective and then propose a comprehensive model of L1 and L2 idiom processing that accounts for these results. In addition, we discuss some possible limitations of this research, its theoretical, methodological and pedagogical implications, as well as new, promising ideas for future avenues of research.

## Chapter 2

# **Normative Data of Dutch Idiomatic Expressions: Subjective Judgements You can Bank on**

**This chapter has been reformatted and slightly modified from:**  
Hubers, F., Cucchiarini, C., Strik, H., & Dijkstra, T. (2019). Normative Data of Dutch Idiomatic Expressions: Subjective Judgments You Can Bank on. *Frontiers in Psychology, 10*, 1-15.

**Abstract**

The processing of idiomatic expressions is a topical issue in empirical research. Various factors have been found to influence idiom processing, such as idiom familiarity and idiom transparency. Information on these variables is usually obtained through norming studies. Studies investigating the effect of various properties on idiom processing have led to ambiguous results. This may be due to the variability of operationalizations of the idiom properties across norming studies, which in turn may affect the reliability of the subjective judgements. However, not all studies that collected normative data on idiomatic expressions investigated their reliability, and studies that *did* address the reliability of subjective ratings used various measures and produced mixed results.

In this study, we investigated the reliability of subjective judgements, the relation between subjective and objective idiom frequency, and the impact of these dimensions on the participants' receptive idiom knowledge by collecting normative data of five subjective idiom properties (Frequency of Exposure, Meaning Familiarity, Frequency of Usage, Transparency, and Imageability) from 390 native speakers and objective corpus frequency for 374 Dutch idiomatic expressions. For reliability, we compared measures calculated in previous studies, with the D-coefficient, a metric taken from Generalizability Theory.

High reliability was found for all subjective dimensions. One reliability metric, Krippendorff's alpha, generally produced lower values, while similar values were obtained for three other measures (Cronbach's alpha, Intraclass Correlation Coefficient, and the D-coefficient). Advantages of the D-coefficient are that it can be applied to unbalanced research designs, and to estimate the minimum number of raters required to obtain reliable ratings. Slightly higher coefficients were observed for so-called experience-based dimensions (Frequency of Exposure, Meaning Familiarity, and Frequency of Usage) than for content-based dimensions (Transparency and Imageability). In addition, fewer raters were required to obtain reliable ratings for the experience-based dimensions. Subjective and objective frequency appeared to be

poorly correlated, while all subjective idiom properties and objective frequency turned out to affect receptive idiom knowledge. Meaning Familiarity, Subjective and Objective Frequency of Exposure, Frequency of Usage, and Transparency positively contributed to idiom knowledge, while a negative effect was found for Imageability. We discuss these relationships in more detail, and give methodological recommendations with respect to the procedures and the measure to calculate reliability.

### **Acknowledgements**

We are grateful to Wendy van Ginkel, who contributed to building the database and setting up the rating study, and to Roeland van Hout for his guidance with respect to the statistics and Generalizability Theory in particular. This work is part of the research program Free Competition in the Humanities with project number 360-70-510 NWO ISLA, which is financed by the Dutch Research Council (NWO).

## 2.1 Introduction

There is a long-standing tradition of research investigating the processing of idiomatic expressions. Assuming that such expressions are stored as chunks with dedicated abstract representations, an analysis of their processing can teach us a lot about how sentence context interacts with the recognition of isolated target words. Unfortunately, large numbers of variables affect idiomatic processing, like familiarity, transparency, and imageability (Cacciari & Glucksberg, 1995; Carrol et al., 2017; Cieślicka, 2006, 2013; Libben & Titone, 2008; Steinel et al., 2007). Studies investigating idiom properties have led to ambiguous results. As an example, Libben and Titone (2008) conducted a series of behavioral experiments on idiom processing and examined the effect of idiom properties on response times. In one of the experiments they found a positive effect of transparency, whereas in another experiment this effect turned out to be absent. A study by Steinel and colleagues (2007) reported a positive effect of imageability, whereas Cacciari and Glucksberg (1995) found that imageability negatively affected idiom processing.

The equivocal results are not only due to the large number of variables and context-sensitivity. Data on idiom properties are usually obtained through norming studies in which subjective judgements are collected for various properties of idiomatic expressions. A consideration of various idiom norming studies shows that the operationalization of relevant variables differs across studies (Beck & Weber, 2016a, 2016b; Bonin et al., 2013; Carrol et al., 2017; Libben & Titone, 2008; Nordmann, Cleland, & Bull, 2014; Nordmann & Jambazova, 2017; Tabossi et al., 2011). For example, in their norming study, Libben and Titone (2008) define familiarity as the extent to which participants have seen, heard or used the idiom, whereas Carrol et al. (2017) use familiarity to refer to the extent participants are familiar with the idiom. Obviously, we will only be able to make some progress in this research domain when relevant dimensions are identified and well-defined. In addition, it is of paramount importance for an appropriate interpretation of the collected data that the dimensions in question are measured reliably. Reliability is “the extent to which measuring instruments (raters) covary, i.e. give relative

values which are correlated” (Rietveld & van Hout, 1993, p. 188). Moreover, reliability needs to be checked before subjective data in subsequent research can be trusted.

However, not all studies that presented normative data on idiomatic expressions investigated their reliability. Studies that did address reliability or agreement of subjective ratings (Bonin et al., 2013; Citron et al., 2016; Nordmann et al., 2014; Nordmann & Jambazova, 2017; Tabossi et al., 2011; Titone & Connine, 1994) employed a whole range of definitions of idiom properties and data collection methods and calculated different measures of reliability. As a result, some studies reported high reliability (Bonin et al., 2013; Citron et al., 2016), whereas others observed low reliability (Nordmann et al., 2014; Nordmann & Jambazova, 2017; Tabossi et al., 2011; Titone & Connine, 1994).

The goal of the present paper is threefold. First, we investigate the reliability of subjective judgements. To this end, we will obtain judgements of various properties of idiomatic expressions that have been prominent in the idiom literature (Frequency of Exposure, Frequency of Usage, Meaning Familiarity, Imageability, and Transparency), and propose the Dependability or D-coefficient (Brennan, 2001) as a measure of reliability, which is relatively unknown in this field of study. Second, we study the relationship between subjective ratings of frequency of exposure and objective ratings of frequency as obtained from corpora. Third, we include an objective measure of idiom knowledge based on meaning recognition for investigating how idiom properties characterized by reliable subjective ratings affect idiom knowledge.

To address these issues, the paper is organized as follows. First, we review previous studies investigating subjective idiom ratings, analyzing how they define the idiom properties under study and apply various reliability measures. Next, we describe how we collected subjective ratings of Dutch idioms for the properties mentioned above in a group of Dutch participants. The ratings are used to calculate different measures of reliability, including the Dependability or D-coefficient. We also assess to what extent reliably measured idiom properties are interrelated and how they affect participants’ knowledge of Dutch idioms. Finally, we discuss our results in relation to those of previous studies and

give some methodological recommendations, proposing the D-coefficient for use in future studies.

### **2.1.1 Properties of idiomatic expressions**

The idiom properties familiarity, transparency, and imageability play a central role in the idiom literature and have been operationalized and defined differently across studies. Familiarity is known to influence idiom processing and is therefore often studied in this type of research. Familiarity has often been defined as “subjective frequency”, indicating how often a given word or idiom is encountered by a speaker (Beck & Weber, 2016a; Gernsbacher, 1984; Titone & Connine, 1994). According to some other authors familiarity “also incorporates how well a meaning is known or understood” (Nordmann et al., 2014, p. 88). Some authors have explicitly addressed this dimension of familiarity by asking subjects to what extent they know the idiom (Cieślicka, 2013) or to indicate how meaningful they find an expression to be (Tabossi et al., 2011). This has also been viewed as a measure of knowledge, albeit one based on subjective self-report.

On closer examination, the terms idiom frequency, familiarity, and knowledge can be taken to refer to distinct, but partially overlapping dimensions. With respect to frequency, a distinction can be drawn between subjective and objective frequency of exposure. The first one could refer to perceived frequency, or the intuition a speaker has of having come across a given expression, while objective frequency can refer to frequency as measured from corpora. However, collecting such objective frequency data for idiomatic expressions is not trivial. First, because it is not immediately clear from which corpus they should be collected, and we know that frequency data are going to vary depending on the corpus used (Gries & Ellis, 2015). Second, owing to the flexible nature of idiomatic expressions, it can be challenging to collect objective idiom frequency from corpora.

Similarly, with respect to familiarity/knowledge, we can discern a subjective variant that indicates to what extent a speaker thinks (s)he is familiar with the meaning of the expression, and a more objective one that indicates to what extent a speaker really knows the expression (subjective and objective meaning familiarity). An additional dimension



may be distinguished that refers to the extent to which speakers use the idiomatic expression themselves, i.e., self-reported frequency of usage. This dimension has not been included in previous studies, but it may be interesting to investigate in the framework of studies on idiom knowledge and idiom production. Therefore, it will be included in the present study (see below).

Imageability, defined as the extent to which a word, or an idiom for that matter, can be associated with a specific image, has been shown to facilitate learning (Paivio, Yuille, & Smythe, 1966). This effect could be a consequence of an additional form of coding beyond verbal coding (Paivio, 1969, p. 257). However, studies on idiom imageability provide rather mixed results. Research on idiom acquisition has indeed shown that imageability has a positive effect on idiom learning (Boers, Lindstromberg, Littlemore, Stengers, & Eyckmans, 2008; Steinel et al., 2007), but Cacciari and Glucksberg (1995) reported a negative effect of imageability on idiom processing. They found that participants more often depict the literal meaning than the figurative one. In addition, Carrol et al. (2017) did not find a clear effect of imageability. However, in their study imageability scores were not obtained from the subjects involved in the study, but were extrapolated from the concreteness norms for single words gathered by Brysbaert, Warriner, and Kuperman (2014).

Transparency is an important property of idiomatic expressions that is often included in idiom processing studies. The exact definition of transparency, however, is not always made explicit and studies have been found to differ in this respect. A clear explanation is provided by Steinel et al. (2007), who refer to the distinction made by Geeraerts (1995) between *syntagmatic transparency* and *paradigmatic transparency*.

*Syntagmatic transparency* is defined as the “one-to-one correspondence between the formal structure of the expression and the structure of its semantic interpretation, in the sense that there exists a systematic correlation between parts of the semantic value of the expression as a whole and the constituent parts of that expression” (Geeraerts, 1995, p. 61). This definition of *syntagmatic transparency* comes close to what other authors refer to as *analyzability* (Cacciari & Glucksberg, 1991; Gibbs & Nayak, 1989; Glucksberg, 1993), *semantic decomposition* (Nunberg, 1979), or *semantic decomposability* “how the

individual meanings of the idiom's component words contribute to the figurative meaning of the phrase" (Titone et al., 2015, p. 173), or "the degree to which individual meanings of an idiom contribute to its overall figurative interpretation" (Cieślicka, 2015, p. 213).

*Paradigmatic transparency*, on the other hand, is defined as the "transparency of the semantic extension that leads from the original meaning of an expression to its transferred reading" (Geeraerts, 1995, p. 61). This corresponds to the definition of transparency adopted by Cieślicka (2015, p. 213): "The extent to which the original metaphorical motivation of an idiomatic phrase can be deduced from its literal analysis". According to this author, the notions transparency and semantic decomposability have often been used interchangeably, while in fact they refer to distinct properties.

Carrol et al. (2017) refer to this distinction by Cieślicka (2015), but eventually opt for another operationalization of transparency and decomposability based on "the stage at which the judgment is being made". In their study, transparency was operationalized as how easily subjects thought they could guess the meaning of the idiom based on the individual words, but without being shown the meaning. Decomposability was defined in the same way, but ratings were obtained later and by showing subjects the correct meaning of the idioms. In between these two questions subjects answered multiple-choice items aimed at testing their knowledge of meaning. While these answers gave the authors information about whether the subjects knew the meanings of the idioms, it is still unclear what the subjects were actually judging when they were asked to rate transparency. Because the actual meaning was not shown, they might have had a different meaning in mind than the correct one, even a meaning that was not included in the multiple-choice items. This complicates the interpretation of the changes in ratings from transparency to decomposability. Further discussion of the relationship between transparency and decomposability (Carrol et al., 2017, p. 17) does not clarify this point. In the current study, we define transparency as *paradigmatic transparency* (Geeraerts, 1995), which is also in line with the definition of Cieślicka (2015).

### **2.1.2 Reliability of subjective judgements**

The variability in various definitions of idiom properties makes it difficult to compare the results of different studies. Moreover, the operationalization of the variables may influence the reliability of the subjective judgements. When the definitions of the dimensions investigated are not unequivocal, participants may be induced to resort to their own interpretations, which in turn may affect the reliability of their ratings. To test this view, we teased apart these interrelated, but distinct dimensions, by proposing different evaluation scales with more precise definitions (see below, and Hubers, van Ginkel, Cucchiari, Dijkstra, and Strik, 2016; van Ginkel, Hubers, Cucchiari, Dijkstra, and Strik, 2016).

Other important elements that may further affect reliability are the research design and the sample size. Most studies collected data using a between-subject design (different groups of participants rated different dimensions of idiomatic expressions), because by using a within-subject design (all participants answered all questions), the ratings on one dimension may be influenced by the ratings on the other dimensions. However, Nordmann and Jambazova (2017) found no effects of study design (within-subjects vs. between-subjects) on idiom ratings. Moreover, “it is important to collect these ratings within subjects, because they can never be independent and should not be treated as such” (Nordmann & Jambazova, 2017, p. 200). In addition, they found that increasing the size of the sample did not improve reliability.

Studies that examined reliability of idiom ratings also differ from each other with respect to the measure of agreement and reliability adopted. This may have consequences for the interpretation of the results concerning reliability. Some studies calculated percentage agreement as a measure of reliability. Titone and Connine (1994), for example, collected normative data for 171 English idioms on various dimensions from groups of 28 to 30 native speakers of English. For familiarity and literalness they employed 7-point scales, but did not measure reliability. In line with Gibbs, Nayak and Cutting (1989), they treated decomposability as a categorical variable distinguishing three categories, and calculated percentage agreement. They concluded: “reliable agreement for compositionality was not found in our study”. Tabossi et al. (2011)

gathered subjective judgements for 245 Italian idiom among 740 Italian native speakers. Groups of at least 40 subjects judged different lists of idioms on various properties on 7-point scales. Reliability was not measured for any of the scales. However, for the compositionality ratings, the percentages of agreement among subjects were compared to those of previous studies by Gibbs and Nayak (1989) and Titone and Connine (1994). In order to do this, the 7-point compositionality scale was converted to a binary variable (decomposable – nondecomposable). As observed by Nordmann et al. (2014), it is unclear what the advantage is of using a 7-point scale if the judgments are then treated as categorical, because in this case relevant information is lost. Tabossi et al. (2011) observed that “for most of the 245 idiomatic expressions judgments were not consistent”, and that “this inconsistency is disturbing as all the studies used the same procedure”. However, the use of percentage agreement is problematic, as this metric does not take chance agreement into account, which makes comparisons across studies difficult. Furthermore, the terms agreement and consistency are used interchangeably here and a measure of agreement for nominal variables, percentage agreement, is used for compositionality ratings on a 7-point scale for which the authors also compute mean and SD values (interval level of measurement).

Another metric that has been used in previous literature is Krippendorff's alpha, an index that is suitable for variables at the nominal, ordinal, interval and ratio level of measurement (Nordmann et al., 2014; Nordmann & Jambazova, 2017). Nordmann et al. (2014) gathered subjective judgments of various idiom properties through 7-point Likert scales from 44 native speakers and 32 non-native speakers of English for 100 English idioms. The reliability values obtained were quite low for both the native and the non-native judgments: They varied between -.02 (familiarity judged by non-natives) and .27 (familiarity judged by natives). Nordmann and Jambazova (2017) describe two rating studies in which reliability (or agreement, the terms are used interchangeably) was measured. The first study employed a larger sample of 160 Bulgarian subjects who rated 90 Bulgarian idioms and a smaller group of 36 English subjects who rated English translations of the Bulgarian idioms. Idiom properties were rated on 7-point Likert scales. Again Krippendorff's alpha was computed and the reliability values appeared to be low in this

case too (between .124 for decomposability and .385 for literality) both for the larger and the smaller groups of subjects. In the second study thirty-two English native speakers were involved in a within-subject rating and 120 took part in between-subject ratings in which four groups of 30 participants rated the same idiom properties as in Study 1. Reliability was low across the board (between .217 for meaning and .332 for familiarity). Inspection of the supplementary materials provided with this paper shows that the authors calculated Krippendorff's alpha for ordinal variables. It is not completely clear whether Likert scales should be treated as ordinal or interval variables, but it is surprising to treat them as interval variables for computing mean and SD values and as ordinal variables for computing reliability.

The low reliability scores obtained in the studies discussed above may be due to the measures used. Both Krippendorff's alpha and percentage agreement are measures of agreement instead of reliability (Tinsley & Brown, 2000). Agreement concerns the absolute values of a set of ratings, and indicates to what extent the values are identical. Reliability, on the other hand, indicates to what extent a set of ratings covary. Reliability can be high even if the absolute values are not identical. Because reliability is based on measures of covariation and correlation, "reliability analysis requires an interval level of measurement" (Rietveld & van Hout, 1993, p. 188). So, the use of agreement or reliability metrics is related to the level of measurement of the variables involved: nominal and ordinal for agreement, and interval for reliability (de Vet, Terwee, Knol, & Bouter, 2006; Rietveld & van Hout, 1993). Moreover, as Rietveld and van Hout (1993) further explain, reliability and agreement measure different aspects of a set of ratings. This point is best illustrated by the discussion presented in Nordmann et al. (2014, p. 93) when they present an analogy from essay assessment: Two teachers assign different grades to two essays by the same student, and the grades by the two teachers for each essay are not identical, but they are strongly correlated. This is a typical case in which a reliability measure will return a high value, but an agreement measure a low one. Nordmann et al. (2014, p. 93) suggest that in the case of normative judgments of idiom properties, we are interested in covariation between the raters and correlation between the ratings, and not so much in whether the values of the ratings are identical. It

follows that in these cases we should compute measures of reliability, not of agreement.

Another metric that has been used in previous research and that does measure reliability is the Intraclass Correlation Coefficient (ICC) (Shrout & Fleiss, 1979). Bonin et al. (2013) collected normative data for 305 French idioms from groups of 23 to 30 French native speakers through 5-point scales. The ICC with random effects of both participants and items was used to measure reliability, obtaining values between .81 for age of acquisition and .96 for subjective frequency. The ICC is an appropriate reliability measure for interval variables, and the parameter setting with random effects of both participants and items allows a generalization to raters not included in the sample. Citron et al. (2016) employed 7-point scales to collect subjective judgements of various idiom properties for 624 German idiomatic expressions by 249 native speakers. Reliability was measured through Cronbach's alpha, a particular case of the ICC, obtaining values between .80 for familiarity and .98 for emotional valence. For Cronbach's alpha raters are treated as a fixed factor and items as random. This parameter setting produces the highest values of ICC. The downside is that in this case the results cannot be generalized to raters not included in the sample (see for further details Rietveld & van Hout, 1993).

The ICC with random effects of both participants and items seems to be the most appropriate reliability measure. It calculates reliability, not agreement, and it allows to generalize across raters. However, the choice of a reliability coefficient may also depend on the presence of missing values. The ICC requires a fully crossed design in which all participants rate all items and is unable to handle missing values. Ideally, we would like to apply a coefficient that can take all these factors into account so as to allow comparisons between studies that differ in various respects from each other. The Dependability coefficient (D-coefficient) based on Generalizability Theory (Brennan, 2001) is one such coefficient.

Generalizability Theory is a statistical theory for evaluating the reliability of behavioral measurements, such as object ratings (Brennan, 2001; Shavelson & Webb, 1991, 2006). The metric proposed for measuring reliability in this framework, and that seems particularly suited for subjective ratings of idiom properties, is the Dependability or

D-coefficient. This metric, based on the ICC, takes into account the estimated variance in items and raters, and is also able to account for the variance in other fixed and random factors (Brennan, 2001; Rietveld and van Hout, in prep.). The D-coefficient has considerable advantages, the most important being that it can take into account sources of variance other than items and raters, and that it can handle different research designs. Regarding the latter, in addition to the fully crossed designs in which each rater judges each item (needed to calculate other reliability measures), Generalizability Theory also allows for unbalanced research designs, in which different groups of participants rate different groups of objects (Brennan, 2001). Another advantage of this statistical theory is that it allows for easy calculation of the minimum number of raters required to obtain reliable data (Li et al., 2015; Shavelson & Webb, 2006). Based on the collected ratings, the number of raters, but also the number of items, can be manipulated to see what the consequences would be for the reliability of the data.

### **2.1.3 The present study**

Our literature review indicates that research on the reliability of subjective judgments of idiom properties so far has been limited and has produced mixed results. Analyses of the studies that investigated reliability reveal a variety of procedures and metrics and suggest that the discrepancies in results may be due to the methods and metrics employed. In the present study, we focused on such reliability issues. In addition, for the idiom properties that could be reliably measured, we investigated their relation with objective idiom knowledge. To that end, we collected and analyzed subjective judgments of frequency of exposure, meaning familiarity, frequency of usage, imageability, and transparency of Dutch idiomatic expressions by Dutch native speakers and their scores on a test of objective knowledge of idiom meaning.

Next, we formulated three research questions. First, we wished to know how reliable subjective judgements of various idiom properties actually are. Thus, we computed their reliability for the newly collected data using Generalizability Theory. In line with Bonin et al. (2013), we expected high reliability values for our ratings, combining the suitability of this technique with more precise definitions and operationalizations of

relevant idiom properties. In addition, subjective ratings of frequency of exposure, frequency of usage, and meaning familiarity were expected to be more reliable than ratings of imageability and transparency, because research indicates that these latter two dimensions are generally more difficult to assess than frequency of exposure, and meaning familiarity. This increased difficulty may be due to a difference in the relation to the idiom. The dimensions meaning familiarity, frequency of exposure, and frequency of usage reflect the native speakers' experience with idiomatic expressions. Because formulaic language, which idiomatic expressions are part of, is found to be generally known by native speakers (Erman & Warren, 2000; Pollio et al., 1977; Wray & Perkins, 2000), and their experience with idiomatic expressions is rather comparable, we expected judgements of these experience-based dimensions to show relatively little variation. The dimensions transparency and imageability, which are more closely related to the content words of the idiomatic expressions, are expected to show more variation. Consequently, subjective judgments of content-related dimensions are expected to be less reliable than judgments of experience-based dimensions.

Second, we wondered to what extent subjective idiom frequency, as assessed in our study, is related to objective idiom frequency as measured from corpora. While subjective and objective frequency have been compared for single words and collocations (Siyanova-Chanturia & Spina, 2015), to our knowledge such systematic comparisons have not been conducted for idiomatic expressions. For single words subjective and objective frequencies appeared to be strongly correlated, whereas for collocations a more complex picture emerged (Siyanova-Chanturia & Spina, 2015). Subjective frequency intuitions of high frequency collocations correlated strongly with objective frequency, as taken from corpora. For medium and low frequency collocations the subjective frequency judgements and objective frequency correlated poorly. As mentioned above, collecting objective frequency data for idiomatic expressions is difficult for a number of reasons related to the choice of the corpus from which the data should be obtained and the flexible nature of idiomatic expressions. We decided to collect this information from the SoNaR corpus (Oostdijk, Reynaert, Hoste, & Schuurman, 2013), a corpus of written Dutch consisting of 500 million words. We opted for this large



corpus to increase the chance of finding idiomatic expressions, given that individual idioms are relatively infrequent. Previous research has shown that subjective frequency of idiomatic expressions is generally relatively high in native speakers (e.g., Beck & Weber, 2016b; Bonin et al., 2013). As to objective frequency, there are indications that while idiomatic expressions as a general phenomenon are frequent, individual idioms are rather infrequent (Ellis, 2012). Based on these findings we expect correlations between subjective and objective idiom frequency to be low.

Third, we were interested to know how different subjective idiom properties and objective idiom frequency are in fact related to objectively assessed idiom knowledge. To answer this question, we reviewed the psycholinguistic literature. Many studies on idiom processing investigated the role of idiom properties in processing (e.g. Cacciari and Tabossi, 1988; Cieślicka, 2013; Gibbs et al., 1989; Libben and Titone, 2008; Titone and Libben, 2014). Only two studies, however, sought to identify idiom properties that are important predictors of offline comprehension measures such as idiom knowledge and subjective familiarity.

Carrol et al. (2017) examined the role of familiarity, and transparency in correctly identifying the meaning of English idiomatic expressions in a multiple-choice question. Familiarity was operationalized as the extent to which participants were familiar with the idiom. Transparency was operationalized as the extent to which participants were able to guess the meaning of the phrase based on the individual words. Carrol et al. (2017) found that Familiarity was a good predictor of objective idiom knowledge, whereas transparency was not found to contribute to idiom knowledge.

Libben and Titone (2008) investigated the impact of idiom properties on the meaningfulness of English idiomatic expressions. Meaningfulness, operationalized as the extent to which participants considered the phrase to be meaningful, can be seen as an indirect and subjective measure of idiom knowledge. In a regression analysis on the aggregated data, the authors examined to what extent familiarity, semantic decomposability, literal plausibility, noun frequency, and verb frequency influenced the meaningfulness ratings. Familiarity was operationalized as what we would define as frequency: the extent to

which the participant has seen, heard or used the idiom. In line with our terminology, we use the term frequency of exposure instead. Frequency of exposure turned out to be an important predictor. The more frequent an idiomatic expression in daily life, the more familiar participants judged this expression to be. Semantic decomposability turned out to be important for infrequent idiomatic expressions only. If an infrequent idiomatic expression was semantically decomposable, people indicated to be more familiar with the idiom, as compared to if the idiom was semantically non-decomposable. The other factors included in the analysis did not significantly influence meaningfulness ratings of English idiomatic expressions.

Both reviewed studies investigated the impact of idiom properties on idiom knowledge. However, Carrol et al. (2017) only examined the effect of familiarity and transparency on idiom knowledge and the operationalization of familiarity was imprecise. Participants could have assessed familiarity with respect to idiom meaning or form. This makes it difficult to interpret the observed positive effect of familiarity on idiom knowledge. Libben and Titone (2008) did investigate the effect of more idiom properties on idiom knowledge, but they assessed idiom knowledge indirectly and subjectively. This assessment shows whether people think they know the meaning of an idiomatic expression, but does not directly tap into the participant's actual idiom knowledge. To investigate how idiom properties influence idiom knowledge, this should be assessed objectively, allowing comparisons between offline (rating) and online comprehension (reaction time) data.

In our study, we investigated the effect of subjective idiom properties and objective idiom frequency on idiom knowledge in more detail. We obtained objective frequency data from a large corpus of written Dutch (Oostdijk et al., 2013) and assessed objective idiom knowledge through multiple-choice questions about the meaning of Dutch idiomatic expressions. We examined more subjective idiom properties than in Carrol et al. (2017), and distinguished three operationalizations of general familiarity: subjective meaning familiarity, subjective frequency of exposure, and subjective frequency of usage. Subjective familiarity is associated with the meaning of the idiom. Subjective frequency is defined as the idiom's occurrence in daily life

(familiarity with the form), and Subjective usage is the extent to which participants indicate to actively use the idiomatic expression themselves. For readability's sake we try to limit the use of the term subjective and opt for the labels Familiarity and Usage, but we maintain Subjective Frequency as opposed to Objective Frequency. We also included Transparency and Imageability in our analysis as predictors of Objective Idiom Knowledge.

Finally, we explored whether the measurements obtained through more precise operationalizations of general familiarity each uniquely contribute to objective idiom knowledge, and how they interact with other idiom properties, such as Transparency and Imageability. Based on the literature, we expected Familiarity, Frequency and Transparency to have a positive effect on Objective Idiom Knowledge (Carrol et al., 2017; Libben & Titone, 2008). As to the effect of Imageability, previous research has been inconclusive. Earlier studies found positive effect of Imageability on idiom learning (Boers et al., 2008; Steinel et al., 2007), whereas it was found to negatively affect idiom processing (Cacciari & Glucksberg, 1995). Objective idiom frequency has not been studied before in this connection. However, other research findings lead to us to assume that objective frequency should have a positive effect on Objective Idiom Knowledge, albeit a less strong one than Subjective Frequency given that the latter is based on individual experience of the same participant.

## **2.2 Materials and method**

### **2.2.1 Participants**

In total, 390 native speakers of Dutch, mainly university students, participated in the rating study (350 female participants and 40 males). Their age varied between 18 and 30 ( $M = 20.4$ ,  $SD = 1.5$ ) and about 98% of them were highly educated. This study was ethically assessed and approved by the Ethics Assessment Committee (EAC) of the Faculty of Arts of Radboud University Nijmegen (number 3382).

### 2.2.2 Materials

We selected 374 Dutch idiomatic expressions and their appropriate meaning based on Dutch dictionaries (e.g., Boon and Hendrickx, 2017; Slot Webcommerce BV, 2017; Stoett, 1925), online idiom lists (Genootschap OnzeTaal, 2017), and our own knowledge and experience. We adjusted these meanings in such a way that they did not contain other idiomatic expressions. For example, to explain the Dutch expression *ergens mee voor de draad komen*, which means “to finally say something”, the dictionary uses another idiom *ergens mee voor de dag komen*. This expressions conveys the same meaning as the expression *ergens mee voor de draad komen*. Therefore, we formulated the meaning in another way without using an idiomatic expression: *iets vertellen* (“to tell something”). The database with the idiomatic expressions and the aggregated results is available in a repository (Hubers et al., 2018).

*Objective idiom frequency.* We collected objective idiom frequency information from the SoNaR corpus of written Dutch (Oostdijk et al., 2013), consisting of 500 million words. First, we identified one content word per idiom (usually a noun) and extracted all sentences from the corpus containing this content word. For example, we looked for all sentences containing the Dutch word *lamp* “lamp” in the corpus (from the Dutch idiom *tegen de lamp lopen* “to get caught”). Second, we obtained the sentences containing the idiomatic expressions in the subset by means of pattern matching, taking into account different word orders and inflections of the verb.

### 2.2.3 Design and procedure

*Operationalization of variables.* Five subjective properties of idioms were rated on 5-point Likert scales: Subjective Frequency, Subjective Usage, Subjective Familiarity, Subjective Imageability and Subjective Transparency (in the remainder of the paper these properties are referred to as Subjective Frequency, Usage, Familiarity, Imageability, and Transparency, respectively). Subjective Frequency is defined as the relative degree to which a participant indicates to have come across an idiomatic expression in speech or in print (Gernsbacher, 1984; Titone & Connine, 1994). Usage is defined as the frequency with which a subject indicates to have used an idiomatic expression. Familiarity is here

conceived of as how well a speaker says to know the meaning of an idiom (Nordmann et al., 2014, p. 88). In line with Steinel et al. (2007), and Boers et al. (2008), Imageability is defined as the extent to which an idiom can evoke an image. This image could be based on the literal or the figurative meaning. Transparency is interpreted in line with (Cieślicka, 2015, p. 213) and paradigmatic transparency (Geeraerts, 1995, p. 61), i.e. the degree to which the semantic value of the entire expression can be understood in terms of the semantic values of its constituting words (Steinel et al., 2007). We also measured knowledge of idiom meaning through an objective multiple-choice test.

*Questionnaire.* The rating study was conducted online through the Qualtrics platform (Qualtrics, 2005). The participants filled in a background questionnaire with questions about gender, year of birth, place of residence, mother tongue, level of education, and language background. In the rating study, the participants answered five questions about the idiomatic expressions on 5-point Likert scales (questions 1,2,3, 4, and 7), one open question (question 5) and one multiple-choice item (question 6).

1. Subjective Frequency: How often have you heard or read this expression? (1. very rarely – 5. very often)
2. Usage: How often have you used this expression yourself? (1. very rarely – 5. very often)
3. Familiarity: How familiar are you with the meaning of this expression? (1. completely unfamiliar – 5. completely familiar)
4. Imageability: How easily can you form an image of this expression? (1. very hard – 5. very easily)
5. Objective Idiom Knowledge (recall): What does this idiomatic expression mean? (open question, not further analyzed in this study)
6. Objective Idiom Knowledge (recognition): Which definition is the correct one? (multiple-choice question: 4 alternatives)
7. Transparency: How clear is the meaning of this expression based on the individual words in the expression? (1. very unclear – 5. very clear)

In line with Titone & Libben (2008) and Bonin et al. (2013), we decided to use a 5-point scale. We opted for a 5-point scale instead of a 7-point scale, because we expected this scale to be sensitive enough to properly reflect people's intuitions about the idiom properties under study.

Since Nordmann and Jambazova (2017) did not find any effects of study design (within-subjects vs. between-subjects) on idiom ratings, we adopted a within-subject design in which all participants answered all questions. This way we take into account the relations between the idiom properties within the individual.

The idiomatic expressions were randomly divided over 15 experimental lists consisting of 25 idiomatic expressions. Every idiomatic expression occurred in only one list. Each participant rated one list of 25 idiomatic expressions and before doing this they rated 2 idiomatic expressions in a practice session in which the questions and the labels of the extreme points of the Likert scales were explained. As a form of calibration, examples were provided of idiomatic expressions representing the extreme values. Participants were not able to go back to their answers on previous questions. On average, the participants completed the rating study in 30 minutes.

#### **2.2.4 Data analysis**

We calculated the mean ratings and standard deviations for all dimensions of each Dutch idiomatic expression. The average Objective idiom knowledge and its standard deviation were calculated based on the proportions correct on the multiple-choice question. To obtain a general overview of the data, we computed the correlations of these dimensions based on the individual data.

To gain insight into the potential differences between reliability measures employed in previous research, we calculated Krippendorff's alpha, Cronbach's alpha, and the Intraclass Correlation Coefficient (ICC) for the data on the different idiom properties obtained in the different experimental lists in our study. These measures were calculated using the 'rel' package (Lo Martire, 2017) in R, version 3.4.0 (R Development Core Team, 2008), and were averaged across lists. We also computed the Dependability coefficient using the 'gtheory' package (Moore, 2016), both averaged across lists and based on the dataset as a whole. The ICC was

calculated for the mean ratings with the parameters ‘two-way’, and ‘absolute agreement’, indicating random effects for participants and items. We refer to this specific instance of the ICC as  $ICC(2,k)$  (Shrout & Fleiss, 1979). To answer the research question on reliability, we compared the D-coefficients based on the dataset as a whole of the different idiom properties, and we calculated the minimum number of raters required to obtain reliable data.

Based on the outcomes of the reliability analyses, we performed logistic mixed effects regression analyses to answer our second research question about the contribution of the different subjective idiom properties to Objective idiom knowledge. These analyses were conducted in the statistical software package ‘R’ version 3.4.0 (R Development Core Team, 2008), and the R packages ‘lme4’ (Bates, Mächler, Bolker, & Walker, 2015), ‘lmerTest’ (Kuznetsova, Brockhoff, & Christensen, 2017), and ‘effects’ (Fox, 2003) were used. The models were built in a forward manner, starting off with a basic model including a random intercept for participants and fixed effects of the idiom properties under study. Subsequently, we added different predictors (random and fixed factors) one by one to the model based on theory, and examined whether the model fit improved. If this was not the case, we decided not to include this predictor in the model. The final model is reported in this paper.

## 2.3 Results

### 2.3.1 General results

Table 2.1 presents a summary of the ratings. In general, participants seem to be exposed to idiomatic expressions quite frequently (mean = 3.41; SD = 1.39), and use idiomatic expressions to a lesser extent (mean = 2.17; SD = 1.30). Idiom knowledge is quite high (85.48% correct). See the Supplementary Materials for the distribution of the individual ratings for the idiom properties (Figure S1).

Table 2.1  
Mean and SD for Ratings on Idiom Properties and for Performance on Knowledge Question

<b>Idiom property</b>	<b>Mean (SD)</b>
Subjective Frequency (scale 1-5)	3.41 (1.39)
Familiarity (scale 1-5)	3.08 (1.35)
Usage (scale 1-5)	2.17 (1.30)
Transparency (scale 1-5)	3.08 (1.28)
Imageability (scale 1-5)	3.36 (1.33)
Objective idiom knowledge (in %)	85.48 (35.22)

Pearson's correlations were computed between the individual ratings for each idiom on all rating dimensions and the objective measures of idiom frequency and idiom knowledge (presented in Table 2.2). All subjective idiom properties significantly correlated with each other, with high values for Subjective Frequency, Familiarity, and Usage (Pearson's  $r > .65$ ). Transparency showed the highest correlation with Objective Idiom Knowledge (Pearson's  $r = .35$ ). Objective Frequency correlated relatively poorly with the subjective idiom frequency judgements (Pearson's  $r = .20$ ), the other subjective judgment scales (Pearson's  $r < .19$ ), and with Objective Idiom Knowledge (Pearson's  $r = 0.08$ ).

### 2.3.2 Reliability

*Reliability measures per list.* We computed the reliability measures for each list separately. Table 2.3 shows the reliability coefficients averaged over the lists. Both the D-coefficient, Cronbach's alpha, and the ICC(2,k) reflect high reliability for each of the idiom properties (all coefficients  $> .85$ ). The D-coefficient and the ICC(2,k) are identical, and Cronbach's alpha is somewhat higher. However, the reliability as reflected by Krippendorff's alpha is much lower for all properties (all coefficients  $< .41$ ). The ratings on Subjective Frequency, Familiarity and Usage seem to be more reliable than the Transparency and Imageability ratings, as indicated by all reliability measures. For a full overview of the coefficients per list, see the Supplementary Materials, Table S1.



Table 2.2  
Correlation Matrix based on Individual Ratings of Dutch Idiomatic Expressions

	<b>Subj.</b>					<b>Obj. idiom</b>
	<b>Frequency</b>	<b>Familiarity</b>	<b>Usage</b>	<b>Transparency</b>	<b>Imageability</b>	<b>knowledge</b>
Familiarity	.79 *					
Usage	.66 *	.68 *				
Transparency	.28 *	.32 *	.29 *			
Imageability	.35 *	.38 *	.27 *	.34 *		
Obj. idiom knowledge	.30 *	.33 *	.24 *	.35 *	.13 *	
Obj. Frequency	.20*	.19*	.19*	-.02	-.02	.08*

Note. \*  $p < .001$



*Reliability measures on entire dataset.* Table 2.4 shows the D-coefficient for the different idiom properties calculated on the entire dataset, taking into account the nested design. The coefficients based on the full dataset are very similar to the averaged D-coefficients and ICCs presented in Table 2.3. The ratings for each of the idiom properties are highly reliable, but those for Subjective Frequency, Familiarity, and Usage are more reliable than those for Transparency and Imageability.

Table 2.4  
D-coefficient for each Idiom Property based on the Full Dataset

<b>Idiom property</b>	<b>D-coefficient</b>
Subjective Frequency	.947
Familiarity	.946
Usage	.937
Transparency	.872
Imageability	.888

*Reliability as a function of the number of raters.* The advantage of Generalizability Theory is that a reliability coefficient can be computed for every number of ratings based on the variance components estimated on the basis of the current data. Figure 2.1 shows the increase in reliability as a function of the number of participants in the rating study. The idiom properties Familiarity, Subjective Frequency, and Usage seem to require fewer raters to collect reliable data as compared to Transparency and Imageability. To obtain highly reliable ratings (D-coefficient > .85) for Familiarity, Frequency, and Usage approximately 10 participants should be recruited. For Imageability and Transparency about 20 people are needed to obtain equally reliable data.

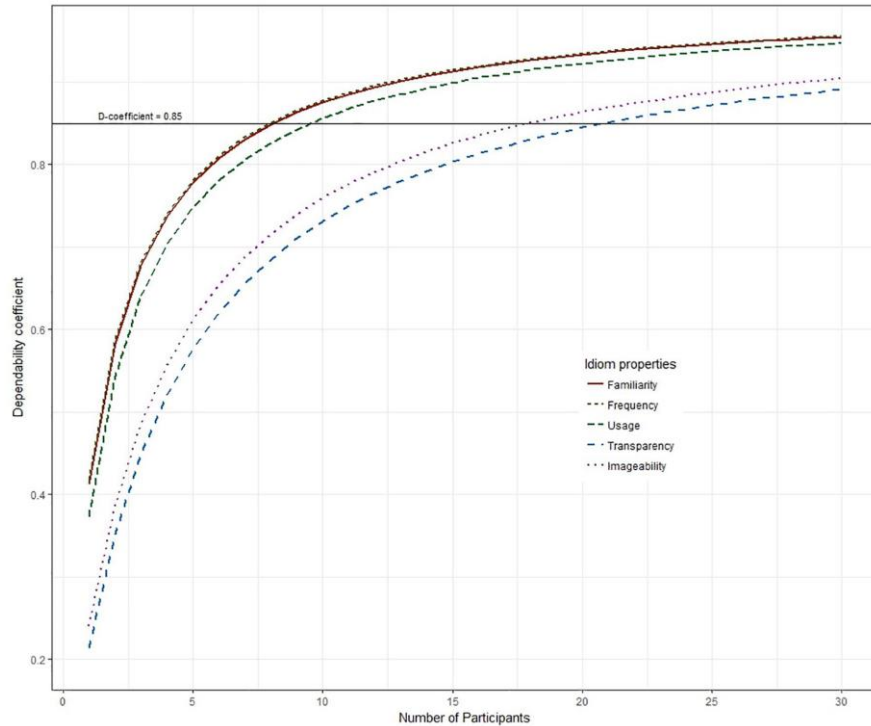


Figure 2.1 Dependability coefficient for a given number of raters.

### 2.3.3 Impact of idiom properties on objective idiom knowledge

To examine which factors influence the participants' knowledge of idiomatic expressions separately and in combination, we conducted a logistic mixed effects regression analysis. The multiple-choice question on idiom knowledge was converted into a binary variable expressing whether the multiple-choice question was answered correctly or not. This binary variable was used as the dependent variable in the regression analysis.

In our final model we included the following predictors as fixed effects: (1) Familiarity, (2) Transparency, (3) Imageability, (4) Subjective Frequency, (5) Usage, (6) Objective Frequency, and the interactions (7) Familiarity x Transparency, and (8) Familiarity x Transparency. All predictors were centered to account for multicollinearity, and Objective Frequency was log-transformed.

In addition, we included Idioms (random intercept only), and Participants as a random effect (random intercept and random slope of Imageability). The model is presented in Table 2.5. The variables included in the model were not collinear (see Variance Inflation Factors in the Supplementary Materials, Table S2).

Familiarity has a positive effect on idiom knowledge ( $\beta = 0.45$ ,  $SE = 0.07$ ,  $p < .001$ ). We also observed a positive effect of Transparency ( $\beta = 0.88$ ,  $SE = 0.05$ ,  $p < .001$ ), Subjective Frequency ( $\beta = 0.14$ ,  $SE = 0.05$ ,  $p < .01$ ), Usage ( $\beta = 0.15$ ,  $SE = 0.06$ ,  $p < .05$ ) and Objective Frequency ( $\beta = 0.16$ ,  $SE = 0.06$ ,  $p < .01$ ) on idiom knowledge. Furthermore, we found a negative effect of Imageability ( $\beta = -0.23$ ,  $SE = 0.05$ ,  $p < .001$ ). The better people are able to form an image of the idiomatic expression, the worse their performance on the multiple-choice question. In addition, we observed a significant interaction of Familiarity and Transparency ( $\beta = -0.07$ ,  $SE = 0.03$ ,  $p < .05$ ) (see the left panel in Figure 2.2). The effect of Transparency on idiom knowledge is larger for idiomatic expressions that are not so familiar as compared to idiomatic expressions that are judged to be highly familiar. This is indicated by the steeper line for unfamiliar idioms than for familiar idioms. Familiarity and Imageability also significantly interact ( $\beta = -0.07$ ,  $SE = 0.03$ ,  $p < .05$ ), see right panel of Figure 2.2. The more familiar participants are with the meaning of the idiomatic expression, the larger the negative effect of Imageability on idiom knowledge. This is indicated by the steeper lines for the familiar idioms than for unfamiliar idioms in the right panel of Figure 2.2.

Table 2.5  
Regression Model with Answer Correct as the Dependent Variable

<b>Fixed effects</b>	<b>Beta</b>	<b>Std. Error</b>	<b>z value</b>
(Intercept)	2.9748	0.1979	15.032 ***
Familiarity	0.4539	0.0647	7.016 ***
Transparency	0.8830	0.0511	17.268 ***
Imageability	-0.2245	0.0496	-4.522 ***
Subjective Frequency	0.1363	0.0517	2.639 **
Usage	0.1531	0.0633	2.418 *
Objective Frequency	0.1623	0.0610	2.660 **
Familiarity x Transparency	-0.0742	0.0295	-2.518 *
Familiarity x Imageability	-0.0687	0.0275	-2.500 *
<b>Random effects</b>	<b>Variance</b>	<b>Std. Deviation</b>	<b>Corr</b>
Idiom	Intercept	2.1537	1.4676
Participant	Intercept	0.6901	0.8307
	Imageability	0.0684	0.2616
			0.36

Note. \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

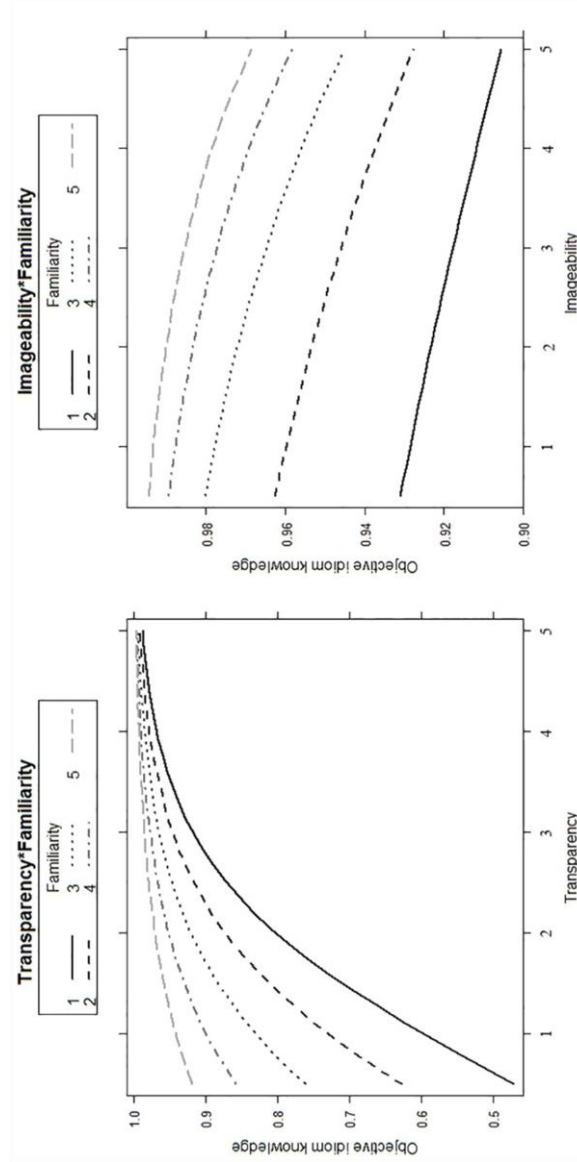


Figure 2.2 The effect of Transparency (A) and Imageability (B) on idiom knowledge for idiomatic expressions with different Familiarity ratings. The variables Transparency, Imageability and Familiarity are rated on a scale from 1 (completely not transparent, imageable, familiar resp.) to 5 (very transparent, imageable, familiar resp.).

## 2.4 Discussion

This study is the first to provide subjective ratings on various dimensions of Dutch idiomatic expressions by native speakers. In order to increase the chances of obtaining an informative picture and reliable ratings, we adopted a more detailed operationalization of familiarity than was employed in previous studies. We found that native speakers indicated being quite familiar with the meaning of idiomatic expressions, to frequently encounter idiomatic expressions in daily life, but to use them less often than that they encounter them. On average, the participants in our study judged the idiomatic expressions to be transparent, and apparently managed to easily form an image of these idiomatic expressions. The objective test of idiom recognition revealed that in general the idiom meanings are well-known. In addition, all subjective idiom properties positively correlated with each other and with Objective Idiom Knowledge.

Despite the difficulties in comparing results across studies because of different operationalizations of the same variables, our findings are in line with those obtained in norming studies on languages other than Dutch. High native speaker ratings of Frequency, and Familiarity are also found for English (Nordmann et al., 2014; Nordmann & Jambazova, 2017), French (Bonin et al., 2013), German (Citron et al., 2016), Italian (Tabossi et al., 2011), and Bulgarian (Nordmann & Jambazova, 2017). Transparency scores are quite comparable to those obtained in other studies (Bonin et al., 2013; Carrol et al., 2017; Citron et al., 2016).

### 2.4.1 Reliability of subjective idiom properties

Many studies have collected normative data on idiomatic expressions and used these as a basis for psycholinguistic research. However, the majority of these studies did not examine whether the collected subjective ratings were in fact reliable (e.g. Carrol et al., 2017; Libben & Titone, 2008). Norming studies that calculated reliability used different metrics and obtained mixed results. Some studies reported poor reliability (Nordmann et al., 2014; Nordmann & Jambazova, 2017), whereas others found high reliability (Bonin et al., 2013; Citron et al., 2016). In this study



we investigated the reliability of judgments of Dutch idiomatic expressions in more detail. We tried to operationalize our variables more precisely than in previous studies, which was expected to increase reliability. Furthermore, to decide which metric to use to assess reliability, we took into account specific aspects of the research design and the level of measurement of the variables. This led us to propose a metric that can be calculated based on the whole dataset and that is relatively unknown in this field of study, the D-coefficient. In contrast to the metrics used in previous studies, this measure can handle unbalanced research designs and missing data. By using this metric, we were also able to assess the minimum number of raters per dimension that are required to obtain reliable data. To show how adopting a metric that is less suitable for the research design can affect reliability, we also calculated the different metrics used in previous studies for our newly collected data.

We calculated different metrics for the individual lists of idiomatic expressions and found that almost all metrics reflect high reliability, except Krippendorff's alpha. We obtained identical values for the ICC(2,k) and the D-coefficient, and slightly higher values (for all dimensions) for Cronbach's alpha. The results of the metrics calculated based on the individual experimental lists show that it is important to use the appropriate metric. As we mentioned above, the ICC(2,k) or the D-coefficient should preferably be used. Krippendorff's alpha is rather a measure of agreement than of reliability, which explains the lower values. Finally, Cronbach's alpha does not consider raters as a random factor, which results in higher reliability scores.

The D-coefficients based on the dataset as a whole, were very similar to the D-coefficients averaged across lists. This shows that the lists were carefully constructed and that the factor List explains only a limited amount of variance. This does not mean, however, that we do not have to take into account the variance of the lists, because it could have been an important source of variance. Moreover, the idiom properties Familiarity, Subjective Frequency, and Usage are highly reliable. The reliability coefficients of the idiom properties Transparency and Imageability are slightly lower, although still very high. These results indicate that by precisely operationalizing the dimensions, using appropriate procedures to obtain the measurements, and by using the

appropriate reliability metric, high reliability can be obtained for subjective judgements of idiomatic expressions.

Based on the reliability coefficients, the dimensions can be divided into two groups: the content-related dimensions Transparency and Imageability on the one hand, and the experience-based dimensions Subjective Frequency, Familiarity, and Usage on the other. This division becomes even more apparent if we consider the minimum number of raters that are required to obtain a reliability of .85. For Familiarity, Subjective Frequency, and Usage approximately 10 participants should be recruited to achieve a reliability of .85, whereas for Transparency and Imageability about 20 participants are needed. In line with our expectations, judging Transparency and Imageability seems to be more difficult than judging Familiarity, Subjective Frequency, and Usage. The experience-based dimensions Familiarity, Subjective Frequency, and Usage appear to be less susceptible to variation than the content-based dimensions Transparency and Imageability.

#### **2.4.2 Comparison between subjective and objective idiom frequency**

To gain more insight into the dimension frequency of exposure, we investigated the relation between subjectively assessed idiom frequency and objective idiom frequency as collected from a large corpus of written Dutch. In line with findings that idiomatic expressions are relatively infrequent (Ellis, 2012) and our expectations about the correlation, we found that Subjective Frequency indeed correlated relatively poorly with Objective Frequency.

As Siyanova-Chanturia and Spina (2015) suggested with respect to collocations, this may be due to the poor ability of people to judge frequency of exposure for low frequency items. An advantage of our study is that we could also check how Subjective and Objective Frequency relate to idiom knowledge. We did find a high correlation between Subjective Frequency and Objective Idiom Knowledge, whereas the correlation between Objective Frequency and Objective Idiom Knowledge was very low. The latter correlation may have been higher if we would have used a corpus of spoken Dutch instead, but these findings can also indicate that Subjective and Objective Frequency reflect different aspects of idiom

frequency. Subjective Frequency as operationalized in our study is closer to individual experience and, apparently, is a better reflection of idiom knowledge than Objective Frequency as obtained from a large corpus of written Dutch. This is not surprising, since the subjective frequency judgements are collected from the same group of participants as the information on idiom knowledge.

In addition, significant correlations were observed between Subjective Frequency and both Imageability and Transparency, while these idiom properties were not related to Objective Frequency. In line with our argumentation and as suggested by one of our Reviewers, this could also explain why Subjective Frequency correlated more strongly with Objective Idiom Knowledge than Objective Frequency: Apparently Objective Frequency is unrelated to the idiom properties that improve idiom knowledge on their own (e.g., Transparency and Imageability). A more detailed study of objective and subjective idiom frequency, their development in native and non-native speakers, and their impact on idiom knowledge and idiom processing would constitute interesting topics for future research.

### **2.4.3 Relation of subjective idiom properties to objective idiom knowledge**

In order to gain more insight into how idiom properties influence receptive idiom knowledge, we investigated how the different subjective idiom properties and Objective Frequency contribute to Objective Idiom Knowledge. We found that all idiom properties significantly impact idiom knowledge. We broke down general familiarity into three more precise operationalizations (Familiarity, Subjective Frequency, and Usage) to see whether each of them uniquely contributed to Objective Idiom Knowledge and how they interacted with other idiom properties. We expected most idiom properties to positively contribute to objectively assessed idiom knowledge. For Imageability, we did not have strong expectations, due to mixed results in earlier studies.

Familiarity, Subjective Frequency, and Usage were found to have a positive effect on Objective Idiom Knowledge, indicating that the more experience users have with the idiom (experience with the meaning, the form, and with using the idiom), the better their idiom knowledge.

Although these dimensions are strongly correlated, there are no signs of multicollinearity in the regression analysis. This, in combination with the fact that all three predictors turn out to be significant in the regression analysis, implies that there is something specific to each of these dimensions that has a positive effect on Objective Idiom Knowledge. Due to the specific and clear operationalizations of these dimensions, the interpretation of these positive effects is more straightforward than that of the broad operationalization of general familiarity as used by Carrol et al. (2017). Moreover, although the correlation with Objective Idiom Knowledge was low, objectively assessed idiom frequency turned out to positively affect Objective Idiom Knowledge. Adding Objective Frequency to the regression model did not change the effects of other predictors. This suggests Objective Frequency has its own unique added value in predicting idiom knowledge, albeit a medium effect only. This, together with the finding that Subjective and Objective Frequency are poorly correlated, confirms our idea that Subjective and Objective Frequency measure different aspects of frequency of exposure.

Transparency also positively influences Objective Idiom Knowledge and contributes most strongly to idiom knowledge. Transparency turned out to be especially important if participants were not familiar with the meaning of the idiomatic expression. Similarly, Libben and Titone (2008) reported an interaction effect between frequency and semantic decomposability in predicting the meaningfulness of a phrase (subjectively assessed). Here the effect of semantic decomposability was especially strong for infrequent idiomatic expressions. Although this interaction effect is slightly different from the interaction effect of Familiarity and Transparency in our study, the underlying reasoning is similar. If participants indicate they are not familiar with an idiomatic expression, they arrive at the meaning of the expression more easily if the idiom is transparent, rather than opaque. This is because in the case of a transparent idiom, the individual words can be used to arrive at the figurative meaning. If participants indicate being familiar with the meaning of the idiomatic expression, Transparency does not affect their performance on the knowledge test, because they know the meaning .

Imageability has a significant, negative impact on idiom knowledge that is stronger for familiar idiomatic expressions than for unfamiliar idiomatic expressions. The direction of the effect is in contrast with earlier studies on idiom learning (Boers et al., 2008; Steinel et al., 2007). Presenting an image of the idiom is found to enhance the link between the form and the meaning of the idiomatic expression (Steinel et al., 2007), resulting in higher learning gains. However, participants may have formed an image of the literal interpretation, rather than of the figurative meaning. This would be in line with Cacciari and Glucksberg (1995), who found that participants more often depict the literal meaning of the idiomatic expression than the figurative meaning. As a result, Imageability negatively affected idiom processing. In the current study, forming an image of the literal interpretation interferes with correctly identifying the idiom's meaning, particularly when participants say that they are familiar with the meaning. If participants are not familiar with the meaning, forming a literal image of the idiom hinders correct recognition of the meaning to a lesser extent.

Being able to form a literal image of the idiom may be related to another idiom property: literal plausibility (Libben & Titone, 2008) or Literality (Beck & Weber, 2016a; Cieślicka, 2006, 2013). This is the extent to which an idiom can be interpreted literally. Libben and Titone (2008) reported a negative effect of literal plausibility on reaction times to idiomatic expressions, an online comprehension measure. Literal plausibility and imageability might be related, because one can relatively easily form an image of idiomatic expressions that are highly literally plausible. This will probably be an image of the literal interpretation, which might interfere with idiom knowledge. For idiomatic expressions that are not literally plausible, the extent to which people are able to form an image may depend on the extent to which they know the meaning of the idiomatic expression. Only if they are familiar with the meaning of the expression, will they be able to form an image of the figurative reading of the idiom. In this latter case, the effect of Imageability would be positive.

## 2.5 Conclusions

Our study addressed subjective judgments by native speakers on idiom properties that are often employed in psycholinguistic research, with the

explicit aim of determining data reliability, the interrelation of the idiom properties and their impact on the participants' idiom knowledge. To this end we performed a comprehensive rating study on Dutch idioms for which the database with idiom properties is now available

Our reliability analysis of subjective judgements by Dutch native speakers with respect to various dimensions of Dutch idiomatic expressions leads us to recommend that future norming studies on idiomatic expressions use the D-coefficient, which is part of Generalizability Theory, as a measure of reliability. The D-coefficient can handle all kinds of research designs and measurement levels, and it allows for generalization across raters. This metric also allows for the assessment of the minimum number of raters that are required to obtain reliable data.

Our study shows that the dimensions can be divided in two groups: experience-based dimensions (Familiarity, Subjective Frequency, and Usage), and content-based dimensions (Transparency, and Imageability). For experience-based dimensions that are carefully operationalized, 10 raters might be sufficient to obtain reliable data, whereas for judgements of the content-based dimensions to be reliable at least 20 participants are required.

Furthermore, the discrepancies between subjective and objective idiom frequency, as observed in this study, suggest that these variables measure different aspects of frequency of exposure. Additional research is necessary to clarify these discrepancies.

Moreover, we found that Transparency, Familiarity, and Imageability most strongly influenced Objective Idiom Knowledge. Imageability negatively influenced idiom knowledge. This negative effect may have been due to a lack of specificity in operationalization, because it is hard to determine whether participants formed an image of the literal or figurative interpretation.

We therefore recommend to researchers that they carefully operationalize idiom properties for their norming studies and assess whether the collected subjective judgements are reliable by using the D-coefficient.

## Chapter 3

### **Second language learner intuitions of idiom properties: What do they tell us about L2 idiom knowledge and acquisition?**

**This chapter has been reformatted and slightly modified from:**  
Hubers, F., Cucchiarini, C., & Strik, H. (submitted). Second language learner intuitions of idiom properties: what do they tell us about L2 idiom knowledge and acquisition?  
*Studies in Second Language Acquisition*

### **Abstract**

The present study investigated intuitions of L2 learners about important properties of L2 idiomatic expressions to gain useful theoretical and methodological insights for research on L2 idiom processing and acquisition. L2 learners' intuitions were checked for reliability, were compared to native speaker intuitions to determine to what extent they differ, whether they are better predictors of L2 idiom knowledge than native speaker intuitions, and to what extent L2 intuitions provide insights into the mechanisms underlying L2 idiom acquisition. To this end, Dutch native speakers and German L2 learners of Dutch rated 110 Dutch idioms on the dimensions frequency of exposure, frequency of use, meaning familiarity, imageability, and transparency. In addition, their idiom knowledge was tested. L2 learner intuitions appear to be highly reliable and reflect L2 idiom knowledge better than those of native speakers. This study shows that L2 learner intuitions about idiom properties are a valuable and reliable source of information that gives more insight into L2 idiom knowledge than native speaker intuitions, and therefore requires more attention. Our results provide important knowledge for theories of L2 idiom processing suggesting that L2 learners essentially apply the same underlying processes as native speakers and that differences between the two groups are mainly caused by differences in degree of exposure to L2.

### **Acknowledgements**

We are grateful to our fellow members of the ISLA team, Wendy van Ginkel, Ton Dijkstra, Louis ten Bosch, and Alexander Wahl for their valuable input and to Ton Dijkstra, Roeland van Hout and Rob Schoonen for their helpful comments on earlier versions of the manuscript. This work is part of the research program Free Competition in the Humanities with project number 360-70-510 NWO ISLA, which is financed by the Dutch Research Council (NWO).



### 3.1 Introduction

Idiomatic expressions like *add fuel to the fire*, *spill the beans*, and *hit two birds with one stone* – usually defined as recurrent sequences of words that convey a figurative meaning (Abel, 2003; Cacciari & Glucksberg, 1991; Titone et al., 2015) – appear to be particularly challenging for second language (L2) learners (Cieślicka, 2006; Conklin & Schmitt, 2008; Ellis et al., 2008; Wray, 2000). While such expressions, and formulaic language in general, are pervasive in native language, they are much less frequent in the L2 of learners (Kecskes, 2007; Pawley & Syder, 1983; Warga, 2005). Even highly proficient L2 learners experience difficulties understanding and using idiomatic expressions (Ellis et al., 2008).

In order to explain why L2 learners find it so difficult to learn idiomatic expressions, and formulaic language in general, Wray has suggested that, in contrast to native speakers, adult L2 learners adopt an essentially “non-formulaic” approach to language learning, primarily focused on individual words (Wray, 2002, pp. 206–209). Other authors have shared this view of the primacy of individual words over formulaic sequences in L2 processing (Cieślicka, 2006, 2015; Kecskes, 2006). Arguments that are often adduced in favor of this more analytic approach are the tradition in L2 teaching of focusing on individual words, the finding that literate learners are aware of words as units in language and have a tendency to focus on those units, and the fact that L2 learners are often already familiar with the individual words contained in formulaic sequences.

On the other hand, Ellis & Wulff (2015) argue that L2 learners adopt a formulaic approach to acquiring formulaic language, memorizing words that often co-occur as chunks, just like native speakers do (Ellis, 2001). However, they also point out that this may be different for idiomatic expressions, since “these are of low frequency, often indeed rare” (Ellis & Wulff, 2015, p. 417).

While the formulaic approach put forward by Ellis (2001) may be more effective than the non-formulaic approach suggested by Wray (2002) for acquiring formulaic language such as collocations, as shown by Durrant and Schmitt (2010), in the case of idiomatic expressions a formulaic, ‘chunking’ approach may be problematic. Words in

collocations are always adjacent, which makes chunking relatively easy and effective. Idioms differ in this respect, as the words in idiomatic expressions may appear adjacently, but this is not necessarily the case, as the word order can be flexible. In addition, word forms may differ because of inflections and conjugations. In the case of idiomatic expressions it may therefore be more difficult for L2 learners to identify whole expressions, recognize them as such and remember them.

Previous research has shown that this is indeed the case, and that this is related to specific properties of idioms such as frequency, familiarity, transparency, imageability, and L1-L2 similarity (García, Cieślicka, & Heredia, 2015; Steinel et al., 2007). Data on these idiom properties are usually obtained by collecting people's intuitions through subjective judgment scales (Bonin et al., 2013; Libben & Titone, 2008; Nordmann et al., 2014; Nordmann & Jambazova, 2017).

In general, such intuitions are collected from native speakers of the language under study, who are considered to be the benchmark. It is generally assumed that L2 learners are not capable of developing reliable intuitions about idiom properties, because they have difficulty in acquiring idioms and have less experience with the language than native speakers. However, investigating L2 learners' intuitions about idiom properties has the potential of providing important insights to L2 acquisition research.

L2 learner intuitions about L2 idiom properties can increase our understanding of the mechanisms underlying L2 idiom acquisition. For instance, findings showing that L2 learners manage to develop reliable intuitions and that these intuitions are related to actual knowledge of idiomatic expressions by L2 learners could be interpreted as evidence in favor of a chunking approach to learning formulaic language, in line with Ellis (2001). On the other hand, findings that intuitions about idiom properties by L2 learners lack reliability and are not related to L2 idiom knowledge could be seen as evidence that L2 learners adopt an essentially non-formulaic approach as suggested by Wray (2002).

So far, relatively few studies have investigated intuitions of idiom properties by L2 learners, and generally for a limited number of idiom properties and with little attention to their reliability. Reliability is considered to be the extent to which raters covary or give relative values

which are correlated (Rietveld & van Hout, 1993, p. 188). If L2 learners turn out to disagree with each other such that their intuitions are inconsistent, the obtained ratings are not reliable and cannot be used in subsequent analyses. Therefore, reliability is an important aspect of subjective ratings that has to be taken into account.

One study that examined reliability reported low reliability scores for both native and non-native intuitions (Nordmann et al., 2014). Other studies did not check the ratings on this aspect (Abel, 2003; Carrol et al., 2017). Titone and Connine (1994) did not explicitly investigate reliability, but their findings do cast doubt on the reliability of subjective ratings of idiom properties by native speakers. Investigating whether L2 learners are at all capable of developing reliable intuitions about idiom properties and whether these intuitions are related to L2 idiom knowledge would provide an important contribution to L2 idiom research.

Considering the importance of L2 learner intuitions about idiom properties for research on L2 idiom processing, its theoretical underpinning and the scarcity of research on this topic, we conducted a comprehensive study of L2 learner intuitions of idiom properties to investigate to what extent (1) L2 learners can develop reliable intuitions about idiom properties, (2) L2 intuitions compare to L1 intuitions about idiom properties, (3) L2 intuitions better reflect receptive L2 idiom knowledge than L1 intuitions, and (4) L2 intuitions can provide insights into the mechanisms underlying L2 idiom acquisition.

In addition to throwing light on the theoretical discussion about L2 idiom acquisition, the present study also attempts to make a methodological contribution to L2 idiom research. Although it might seem reasonable to take native intuitions as the benchmark, because native speakers can be considered as the model the L2 learners are trying to achieve, it is conceivable that native intuitions do not reflect L2 knowledge, but rather provide a distorted picture that is not in line with the intuitions and impressions of L2 learners. Because this could have a biasing effect on the results of research, collecting intuitions directly from L2 learners would then seem to be preferable. Investigating whether these intuitions are reliable and whether they are more informative

about L2 idiom knowledge than native intuitions could contribute to improving the methodology in L2 idiom research.

The paper is organized as follows. First, we discuss the formulaic and non-formulaic approach to L2 language acquisition in more detail. We go on to discuss important idiom properties in idiom research, and review studies that examined the reliability of subjective ratings of idiom properties. We then introduce the current study in which we collected ratings of various idiom properties from native speakers and L2 learners. We go on to calculate the reliability of the L1 and L2 intuitions, compare these, and examine to what extent reliable intuitions can be employed to explain L2 idiom knowledge. Finally, we present our results and discuss them in relation to theories of language learning and acquisition.

### **3.1.1 L2 idiom acquisition**

Ellis (2001) suggests that native speakers learn language based on frequency information and associative learning. Words that often co-occur tend to be memorized as chunks. One of the questions is to what extent L2 learners use a similar mechanism in learning formulaic language and, more specifically, idiomatic expressions. Several studies on L2 formulaic language processing have shown processing advantages for formulaic sequences as opposed to control phrases suggesting that these expressions are stored as chunks in the mental lexicon (Conklin & Schmitt, 2008; Isobe, 2011; Jiang & Nekrasova, 2007). In later work, Ellis & Wulff (2015) argue that L2 learners may indeed adopt a similar formulaic approach to learning formulaic expressions, but they seem to make exceptions for idiomatic expressions, which are less frequent and sometimes peculiar.

As introduced above, Wray (2002) suggests that adult L2 learners rely more on the individual words in language learning as compared to children acquiring their first language. She argues that children acquire L1 formulaic language holistically, since in the earlier stages of language acquisition they are not yet able to segment incoming speech. Only after learning to read and write, do native speakers start to break down chunks. Adult L2 learners have ample experience with reading and writing and therefore tend to analyze the incoming speech signal in smaller lexical units (Wray, 2002, 2008).

The idea that L2 learners rely more on the individual words while learning formulaic language has been embraced also in the domain of L2 idiom processing (Abel, 2003; Cieśllicka, 2006; Giora, 1997; Kecskes, 2006). Abel (2003) examined offline ratings by German learners of English. She found that L2 learners judged non-transparent idioms as transparent, indicating that they believed that the individual word meanings contribute to the figurative meaning even if this is not the case. In an online cross-modal priming experiment, Cieśllicka (2006) observed a processing advantage for literal targets as compared to idiomatic and control targets in Polish speakers of English. She interpreted this in favor of her Literal Salience Model that assumes primacy of literal word meanings over the figurative meaning during L2 idiom processing. Kecskes (2006) argues that salience is determined by familiarity and experience with a given meaning. An idiom's figurative meaning may not be salient at the start, but can become more salient with time and repeated exposure.

### **3.1.2 Idiom properties**

#### *Frequency and familiarity*

The familiarity and frequency of idiomatic expressions are important dimensions in idiom processing research. These terms are often used interchangeably and although there are various ways to define them, a prominent definition of idiom familiarity is the extent to which people indicate to be familiar with (the meaning of) the idiomatic expression (Abel, 2003; Hubers, Cucchiarini, Strik, & Dijkstra, 2019; Hubers et al., 2018; Nordmann et al., 2014), while idiom frequency is often defined as the frequency with which a speaker or listener indicates to have encountered an idiomatic expression (Carrol et al., 2017; Gernsbacher, 1984; Libben & Titone, 2008).

All rating studies that collected L1 and L2 intuitions about idiom familiarity and idiom frequency found higher ratings for native speakers than for L2 learners (Abel, 2003; Carrol et al., 2017; Nordmann et al., 2014). This is not surprising considering that native speakers are exposed to their native language throughout their lives, while L2 learners receive only limited input in their L2.

In addition to subjective idiom frequency, frequency can also be measured objectively, on the basis of corpora. Collecting individual word frequencies from corpora is rather straightforward, but collecting objective idiom frequency data can be a challenge, because of the flexible nature of idiomatic expressions (i.e. different possible word orders and inflections). Only a few studies have compared subjective and objective frequency of units larger than single words (Hubers et al., 2019; Siyanova-Chanturia & Spina, 2015). Hubers et al. (2019) collected subjective frequency judgements of idiomatic expressions by native speakers and found that subjective and objective idiom frequency both appeared to uniquely contribute to idiom knowledge. Siyanova-Chanturia and Spina (2015) obtained subjective frequency judgements from native speakers and L2 learners about collocations, another type of formulaic language, and compared these to objective frequency data obtained from corpora. They reported strong correlations between frequency intuitions of both native speakers and L2 learners and objective frequency for high frequency collocations. For medium and low frequency collocations, no significant correlations were observed, and for the very low frequency collocations only the L1 frequency intuitions correlated with objective frequency. Siyanova-Chanturia and Spina (2015) underline the importance of studying language users' intuitions about the frequency of units that transcend single words, like collocations and other forms of multiword expressions, as evidence accumulates that these are an important component of language, while still little is known about how they are processed by L2 learners.

Given that individual idioms are not particularly frequent and that, consequently, L2 learners are not likely to encounter them often in naturalistic L2 input (Ellis, 2012), an important question is whether L2 learners have enough opportunities for developing intuitions about idiom frequency and familiarity. Over and above the reduced L2 input, an additional factor that might hinder L2 learners developing intuitions about idioms may be their difficulties in noticing formulaic language. L2 learners are more likely to fail to notice formulaic expressions even when they encounter them (Boers & Lindstromberg, 2012; Eyckmans, Boers, & Stengers, 2007; Jones & Haywood, 2004; Peters, 2012). Idiomatic expressions containing familiar words more often go unnoticed by L2

learners than idiomatic expressions containing unfamiliar words (Kim, 2016; Laufer, 1997).

### *Transparency*

Transparency is generally defined as the degree to which the semantic value of the entire expression can be understood in terms of the semantic values of its constituting words (e.g., Steinel et al., 2007) and is often measured by asking native speakers to indicate to what extent they “consider an idiomatic expression as related to its figurative meaning” (Skoufaki, 2008, p. 20). The idiom *spill the beans* is opaque, because the figurative meaning *to reveal a secret* cannot be extracted from the literal interpretation. The expression *to hit two birds with one stone* is transparent, because the figurative meaning (*to solve two problems at once by a single action*) can be extracted from the literal interpretation. Transparent idioms appear to pose fewer problems to L2 learners than opaque ones in terms of idiom production and comprehension (Irujo, 1986b; Skoufaki, 2008; Steinel et al., 2007; Yorio, 1989).

Because idiomatic expressions are imbued with specific linguistic and cultural knowledge (Boers et al., 2004; Kovecses & Szabó, 1996), it is to be expected that L1 and L2 transparency intuitions are different. Boers and Webb (2015) compared transparency intuitions of English idioms by native speakers with those of advanced learners of English, and found that the L1 and L2 intuitions were quite different. Abel (2003) investigated intuitions of semantic decomposability, a concept related to transparency, by L2 learners of English, and concluded that L2 learners tend to rely more on literal meanings than the native speakers in a comparable study of Titone and Connine (1994). In Nordmann et al. (2014) L2 learners judged idioms to be less literal and less decomposable than native speakers, whereas in Carrol et al. (2017), native speakers judged English idioms to be less transparent than non-native speakers.

Researchers in cognitive linguistics (Geeraerts, 1995; Gibbs, Bogdanovich, Sykes, & Barr, 1997) maintain that transparency intuitions are, at least partly, influenced by inherent properties, like conceptual metaphors and encyclopedic knowledge (Skoufaki, 2008). Keysar and Bly (1995) argued that transparency intuitions are not necessarily rooted in the motivation underlying idioms, but emerge because language users

develop explanations for the meanings they have learned to associate with specific idioms. They investigated this by presenting native speakers with scenarios in which an unfamiliar idiom appeared towards the end of the scenario. In a learning phase, the scenarios biased the reader towards (1) the original idiom meaning, (2) the reversed meaning, or (3) an unrelated meaning. For each scenario in the learning phase, participants had to choose the meaning of the italicized idiom out of three alternatives (original meaning, reversed meaning, unrelated meaning). In the test phase, participants were presented with the same idioms, but in an unbiased scenario. These scenarios included a person who encountered an unfamiliar idiom. Participants had to indicate the way this person would interpret the idiom by selecting one of the three alternatives. Keysar and Bly (1999) suggested that native speakers developed explanations for the learned idiom meanings post facto, while alternative, unlearned meanings became less accessible. A similar conclusion was drawn by Malt and Eiter (2004) with respect to L2 learners. However, Skoufaki (2008) challenged Keysar and Bly's view, and ascribed their findings in part to specific features of their experiment (overrepresentation of opaque idioms in their material and a task that pre-empted the use of idiom-inherent properties (Skoufaki, 2008, p. 22). A study by Ramonda (2019) on semantic transparency intuitions of idioms by English native speakers also appeared to contradict the highly arbitrary nature of semantic transparency suggested by Keysar and Bly (1999). In order to gain more insight into the source of transparency intuitions, Skoufaki (2008) presented advanced L2 learners of English with unknown idiomatic expressions, varying along the transparency dimension, and asked them to guess the meaning and provide an interpretation. She found that high-transparency idioms received fewer different interpretations than low-transparency idioms, which led her to propose a hybrid view of idiom transparency, in which not only idiom familiarity or knowledge affect transparency intuitions, but also idiom-inherent features, i.e. the individual words.

The present study systematically compares L1 and L2 transparency intuitions, and investigates these issues making it possible to test different hypotheses. If it is essentially idiom familiarity that drives transparency intuitions, as Keysar and Bly (1995) suggest, then



transparency ratings by native speakers should be higher than those by L2 learners. If, on the other hand, transparency intuitions are also affected by intrinsic idiom properties, as proposed by Skoufaki (2008), then it is possible that L2 learners judge the same idioms to be at least as equally transparent as native speakers do. In other words, similar or higher L2 transparency ratings would suggest that transparency intuitions also have a more objective, idiom-inherent basis and are not only induced by idiom familiarity.

### *Imageability*

Imageability indicates the degree to which an idiom can evoke an image (Cacciari & Glucksberg, 1995; Steinel et al., 2007). Cacciari and Glucksberg (1995) found that in native speakers mental images are usually associated with the literal meaning of idioms rather than with the figurative one. This could imply that the degree to which an image can be formed of an idiom may hamper processing rather than facilitate it.

Research on L2 idiom acquisition has shown that the extent to which idioms can be associated with images has a positive effect on learning the meaning of L2 idioms (Steinel et al., 2007). This is in line with the dual coding hypothesis (Paivio, 1986; Sadoski, 2005), which assumes that cognition occurs in a verbal code for language and a non-verbal code for mental imagery. However, Boers et al. (2008) found that pictorial elucidation was not conducive to better retention of the linguistic form of the idioms.

The present study makes a direct comparison between L1 and L2 imageability intuitions. In addition, by investigating the impact of imageability on idiom knowledge we expect to gain a better understanding of the processes underlying L2 idiom acquisition.

### *Cross-language overlap*

While the idiom properties discussed above are relevant for research on idiom processing in both L1 and L2, L2 studies need to take account of an additional dimension: Idiom translatability or cross-language overlap/similarity (Charteris-Black, 2002; Cieślicka, 2006, 2015, Irujo, 1986a, 1986b, Liontas, 2002, 2015b). Previous research indicates that L2 idiom comprehension and production are affected by cross-language

overlap. Idioms that exist in both L1 and L2 appear to be less difficult (Carrol et al., 2017; Irujo, 1986b; Laufer, 2000), but this seems to hold especially for comprehension, while in production similar idioms may be more problematic because of interference from L1 (Charteris-Black, 2002; Liontas, 2002; Yorio, 1989).

The extent to which L2 idioms exist in the L1 can also influence L2 intuitions about idiom properties. L2 idioms with exact equivalents in L1 appear to be judged as more familiar and more transparent than L2 idioms which do not have identical matches in L1 (Carrol et al., 2017). It is not clear, however, how L2 intuitions are influenced by intermediate levels of cross-language overlap and how cross-language overlap affects the relation between subjective and objective characteristics of L2 idioms.

A more detailed classification that takes account of both form and meaning as proposed by Titone et al. (2015) seems to be required to obtain a clearer understanding of how cross-language overlap affects L2 intuitions of idiom properties. These authors used a scale ranging from 1 to 5 and found that cross-language overlap facilitated idiom processing. The current study examines cross-language overlap and relates this to L2 idiom knowledge.

### **3.1.3 Reliability**

L2 learners are generally less exposed to the L2 and in particular to L2 idioms (Wray, 2002) than native speakers. As a result, they are likely to develop less reliable intuitions about the idiom familiarity, frequency, transparency, and imageability. However, this might be modulated by the proficiency level of the L2 learners, the amount of L2 experience, and their native language.

Nordmann et al. (2014) investigated L1 and L2 intuitions and examined their reliability by collecting ratings of familiarity, meaning, literalness, and decomposability through 7-point Likert scales from 44 native speakers and 32 non-native speakers of English for 100 English idioms. The authors analyzed the reliability of the ratings and concluded that it was low for both L1 and L2 intuitions. The diversity among the non-native speakers' native languages might have caused differences in the ratings that affected the degree of reliability. In a more homogeneous

sample of participants with the same L1 reliability should be higher, although this might seem less plausible given the low reliability values that Nordmann et al. (2014) obtained for native speakers, who constitute a more homogeneous group. The present study will throw more light on this issue.

### **3.1.4 The present study**

The review of previous research on L2 intuitions about idiom properties reveals that a number of important and crucial questions remain unanswered. These concern the reliability of L2 intuitions, the differences between L1 and L2 intuitions, their relevance to theoretical models of idiom acquisition and their possible consequences for subsequent research on L2 idiom processing. Moreover, it is not yet clear how L1 and L2 intuitions and cross-language overlap are related to an objective measure of L2 idiom knowledge.

To investigate these issues, we collected intuitions of frequency, familiarity, usage, transparency, and imageability of Dutch idiomatic expressions from L1 German learners of Dutch L2 and native speakers of Dutch, data on objective frequency of idioms obtained from corpora and objectively assessed meaning recognition as a measure of idiom knowledge. In the remainder of this paper we refer to this specific type of receptive idiom knowledge as L2 idiom knowledge. We addressed the following research questions:

1. Are L2 learners capable of developing reliable intuitions about idiom properties?
2. How do L2 intuitions compare to L1 intuitions about idiom properties?
3. Do L2 intuitions better reflect L2 idiom knowledge than L1 intuitions?
4. Can L2 intuitions provide insights into the mechanisms underlying L2 idiom acquisition?

With respect to the first research question, we hypothesize that L2 learners are capable of developing reliable intuitions about frequency, usage, familiarity, imageability, and transparency, but that these are less

reliable than L1 intuitions, since L2 learners are much less exposed to the target language and culture. Although Nordmann et al. (2014) found that the reliability of L2 intuitions of idiom properties was very low, we hope to increase the chance of obtaining reliable results by adopting a more suitable statistical measure of reliability and more specific questions about the idiom properties under study, since this is an important aspect that can influence the results (Hubers et al., 2019).

With regard to the second research question, we expect that limited exposure to the L2 leads to lower ratings for familiarity, frequency, and usage by L2 learners than by native speakers. More specifically, we hypothesize that L2 learners will indicate that they are less familiar with idioms and that they encounter and use them less frequently than native speakers do.

In relation to transparency, we are interested in comparing the predictions by Keysar and Bly (1995) with those made by Skoufaki (2008). If transparency is mainly influenced by idiom familiarity, as Keysar and Bly (1995) suggest, then we expect higher transparency ratings for native speakers as compared to L2 learners. However, if transparency intuitions are also affected by idiom intrinsic properties, as proposed by Skoufaki (2008), then native speakers are expected not to judge idioms as more transparent than L2 learners do.

As to imageability, the picture is less clear-cut, owing to the sparseness of research findings on this topic. Native speakers tend to associate mental images with the literal meanings of idioms rather than with the figurative ones (Cacciari & Glucksberg, 1995). Given their higher proficiency, they should be more likely to link idioms to images. However, the limited research on the role of imageability in idiom learning suggests that imageability can have a facilitative role (Steinel et al., 2007). We might therefore expect that L2 learners exploit this more than native speakers do.

To gain more insight into the relationship between the L1 and L2 intuitions, we will also check the correlations between these ratings. Intuitions of frequency, familiarity, and usage are more experience-based while intuitions of transparency and imageability are more related to intrinsic properties of the idioms themselves (Hubers et al., 2019). For these reasons we should expect stronger correlations between L1 and L2

intuitions of transparency and imageability, than for those of frequency, familiarity, and usage.

With regard to the third question, our hypothesis is that L2 intuitions are better predictors of L2 idiom knowledge than L1 intuitions. Furthermore, on the basis of the findings from Titone et al. (2015), we expect a positive effect of cross-language overlap on L2 idiom knowledge. Finally, another interesting element in this respect is the role of objective frequency. Research on native speakers by Hubers et al. (2019) suggests that subjective and objective idiom frequency both uniquely contribute to receptive idiom knowledge. It is interesting to see to what extent these two frequency variables, one based on intuitions and one based on corpora, contribute to L2 idiom knowledge.

As to the fourth research question, we expect L2 learners' intuitions about idiom properties to indicate that L2 learners are capable of applying a chunking approach just like native speakers, as suggested by Ellis (2001). However, based on the considerations about idiomatic expressions in Ellis & Wulff (2015), it may well be the case that the results vary depending on the specific properties of the idioms. Opaque idioms, for example, are more likely to be treated as chunks, while L2 learners may rely more on the individual words in transparent idioms.

## 3.2 Methods

### 3.2.1 Participants

*Native speakers.* In total, 26 native speakers of Dutch participated in our study (24 females). They were mainly university students, were on average 22.7 years old, ranging from 19 to 34 ( $SD = 3.2$ ), and obtained an average score of 90.82 ( $SD = 6.07$ ) on the LexTale, a measure of vocabulary knowledge ranging from 0 to 100 (Lemhöfer and Broersma 2012).

*L2 learners.* In total, 26 L2 learners participated in our study (23 females). They were German speakers of Dutch studying or working at a Dutch university. They were between 21 and 32 years old (mean age = 24.76,  $SD = 3.46$ ), and were moderately to highly proficient in Dutch, as assessed by the LexTale (mean = 69.04,  $SD = 11.75$ ) (Lemhöfer & Broersma, 2012). They all started learning Dutch around the age 18 to 20.

See Table 3.1 for an overview of the participant characteristics. This study was ethically assessed and approved by the Ethics Assessment Committee (EAC) of the Faculty of Arts of Radboud University Nijmegen (number 3382).

Table 3.1  
An Overview of the Participant Characteristics per Participant Group.

	<b>Native speakers</b>	<b>L2 learners</b>
Age	22.69 (3.15)	24.76 (3.46)
LexTale (1-100)	90.82 (6.07)	69.04 (11.75)
Self-reported proficiency (scale 1-7)	NA	5.51 (1.11)
#hours a week speaking Dutch outside class or work	NA	9.66 (17.46)
#years learning Dutch	NA	3.65 (2.38)
#years living in the Netherlands	NA	1.45 (2.29)

### 3.2.2 Materials

We selected 110 Dutch idioms from a database consisting of 393 idiomatic expressions rated by native speakers on various idiom properties, such as Familiarity, Transparency, and Imageability (Hubers et al., 2019, 2018). We selected the idioms in such a way that they varied along Familiarity, Transparency, and Imageability, resulting in a plausible reflection of the variation in the full dataset. To design multiple choice items for the knowledge test we created three incorrect alternative meanings that would be plausible if one were not familiar with the idiom. For the idiom *de boeken sluiten* ('to close the books', meaning 'to go bankrupt'), for example, we created the following alternative meanings: 'spend much money', 'take a risk', 'go into retirement'. See Table S3 in the Supplementary Materials for the Dutch idiomatic expressions included in this study.

*Cross-language overlap.* For the 110 Dutch idioms, the degree of similarity between Dutch and German was determined by two bilingual German-Dutch students. They assessed cross-language overlap using a slightly adapted version of the rating system described in Titone et al. (2015). Four levels of overlap were distinguished: (1) The Dutch idiom

does not exist in German (NE), (2) The Dutch idiom does exist in German, but in completely different content words (DW), (3) The Dutch idiom does exist in German and has  $n$  content words in common (nW), (4) The Dutch idiom has a word-to-word correspondent in German (AW). The students individually scored all idioms and subsequently compared their scores. The students assigned different cross-language overlap scores to only 5 idioms. These differences were resolved by discussing these expressions in more detail. In the end all 110 idioms were assigned a cross-language overlap score.

*Objective idiom frequency.* We collected objective idiom frequency information from the SoNaR corpus of written Dutch (Oostdijk et al., 2013), consisting of 500 million words. First, we identified one content word per idiom (usually a noun) and extracted all sentences from the corpus containing this content word. For example, we looked for all sentences containing the Dutch word lamp “lamp” in the corpus (from the Dutch idiom *tegen de lamp lopen* “to get caught”). Second, we obtained the sentences containing the idiomatic expressions in the subset by means of pattern matching, taking into account different word orders and inflections of the verb.

### 3.2.3 Design and Procedure

*Operationalization of the idiom properties.* Five subjective idiom properties were measured through ratings on 5-point Likert scales: Frequency, usage, familiarity, imageability, and transparency. In addition, we measured objective knowledge of idiom meaning through a multiple choice test.

Frequency is defined as the relative degree to which a participant indicates to have come across an idiom in speech or in print (Gernsbacher, 1984; Titone & Connine, 1994). Usage is defined as the frequency with which a subject indicates having used an idiom. Familiarity is conceived of as how well speakers say that they know the meaning of an idiom (Nordmann et al. 2014: 88). In line with Steinel et al. (2007), and Boers et al. (2008), imageability is defined as the extent to which an idiom can evoke an image. This image could be based on the literal or the figurative meaning. Transparency is interpreted as “the extent to which the original metaphorical motivation of an idiomatic

phrase can be deduced from its literal analysis” (Cieślicka, 2015, p. 213), which is in line with paradigmatic transparency as defined by (Geeraerts, 1995, p. 61) as the degree to which the semantic value of the entire expression can be understood in terms of the semantic values of its constituting words (Steinel et al., 2007). Objective idiom knowledge is considered as a measure of participants’ familiarity with the meaning of the idiom, as obtained from a multiple-choice test of meaning recognition.

*Questionnaire.* The rating study was conducted online through the Qualtrics platform (Qualtrics, Provo, UT). The participants started by filling in a background questionnaire. The L2 learners were asked to provide information on their language background. More specifically, we asked for self-reported proficiency in Dutch on speaking, writing, reading and listening (on a scale ranging from 1 to 7), the number of years they had been learning Dutch, and the number of hours a week they used Dutch outside class or work.

In the main part of the rating study the participants had to answer five questions about the idiomatic expressions on 5-point Likert scales (questions 1, 2, 3, 4, and 6), and one multiple choice item (question 5):

1. Frequency: How often have you heard or read this expression?  
(1. very rarely – 5. very often)
2. Usage: How often have you used this expression yourself?  
(1. very rarely – 5. very often)
3. Familiarity: How familiar are you with the meaning of this expression? (1. completely unfamiliar – 5. completely familiar)
4. Imageability: How easily can you form an image of this expression? (1. very hard – 5. very easily)
5. Objective idiom knowledge (recognition): Which definition is the correct one? (multiple choice question: 4 alternatives)
6. Transparency: How clear is the meaning of this expression based on the individual words in the expression?  
(1. very unclear – 5. very clear)

In line with Titone & Libben (2008) and Bonin et al. (2013), we decided to use a 5-point scale. We opted for a 5-point scale instead of a 7-point



scale, because we expected this scale to be sensitive enough to properly reflect people's intuitions about the idiom properties under study.

By using a within-subject design, the ratings on one dimension may be influenced by the ratings on the other dimensions. However, Nordmann and Jambazova (2017) found no effects of study design (within-subjects vs. between-subjects) on idiom ratings. Moreover, "it is important to collect these ratings within subjects, because they can never be independent and should not be treated as such" (Nordmann & Jambazova, 2017, p. 200).

The idioms were organized in four blocks of 27, 28, 28, and 27 expressions respectively. These four sets were chosen based on the L1 ratings, which led to an even distribution of familiar and unfamiliar idioms (according to native speakers) across these four blocks. The order of presentation within blocks was randomized. The participants were allowed to complete the blocks at their own pace. In the first block, both the native speakers and the L2 learners carried out the Dutch version of the LexTale vocabulary test (Lemhöfer & Broersma, 2012), as an indicator of their proficiency in Dutch. It took the participants between 30 and 45 minutes to complete each block.

#### **3.2.4 Data analysis**

In order to address the first research question, we examined the reliability of the L1 and L2 intuitions by calculating the Intraclass Correlation Coefficient (ICC) using the 'rel' package (Lo Martire, 2017) in R, version 3.4.0 (R Development Core Team, 2008). The ICC was calculated for the averaged ratings with the parameters 'two-way', and 'absolute agreement', indicating random effects for participants and items. We also examined the reliability of the objective idiom knowledge test by calculating Cronbach's alpha using the same R package (parameters 'two-way' and 'consistency').

To answer the second research question we then compared the L1 and L2 intuitions by computing the mean ratings and standard deviations for all subjective dimensions for native speakers and L2 learners separately. The proportions correct on the multiple choice question were taken to calculate the average objective idiom knowledge and standard deviation. Pearson's correlations were calculated between

the L1 and L2 intuitions, and between the L1 and L2 idiom knowledge based on aggregated data.

To examine to what extent L1 or L2 intuitions better reflect L2 idiom knowledge (research question 3), we performed logistic mixed effects regression analyses based on the individual data. We used the statistical software package ‘R’, version 3.4.0 (R Development Core Team, 2008), and the R packages ‘lme4’ (Bates et al., 2015), ‘lmerTest’ (Kuznetsova et al., 2017), and ‘effects’ (Fox, 2003) to conduct the analyses. The models were built in a forward manner. Since L1 intuitions are often taken as benchmarks in idiom processing studies, we started off with an initial model including a random intercept for participants and fixed effects of the idiom properties under study as rated by the *native speakers*. Subsequently, we added the ratings based on L2 intuitions (as fixed factors) one by one to the model and examined whether the model fit improved. If this was not the case, the predictor was not included in the model. Next, objective frequency, cross-language overlap, participant characteristics (fixed factors) and potential random factors were added using the same procedure. If the model fit did not improve, the predictor was not included in the model. During this process, we also excluded predictors that did not significantly contribute to the model fit. Both the initial and the final model, based on L1 and L2 intuitions, respectively, are reported in this paper.

### 3.3 Results

#### 3.3.1 Reliability

Table 3.2 shows the ICC for the various idiom properties included in our rating study per participant group. Cronbach’s alpha is presented as a reliability measure of the idiom knowledge test. The L1 and L2 ratings for all dimensions were highly reliable (ICC > 0.9, and ICC > 0.85 respectively), as well as the L1 and L2 performance on the objective idiom knowledge test (Cronbach’s alpha = 0.91 for both groups).

Table 3.2  
ICC for each Idiom Property and Cronbach's Alpha for Objective Idiom Knowledge per Participant Group.

<b>Idiom property</b>	<b>Native speakers</b>	<b>L2 learners</b>
Frequency	0.94	0.88
Familiarity	0.94	0.93
Usage	0.94	0.86
Transparency	0.94	0.90
Imageability	0.91	0.89
Objective idiom knowledge	0.91	0.91

### 3.3.2 Comparison of L1 and L2 intuitions

Table 3.3 presents the mean ratings of the various dimensions and their standard deviations as provided by the Dutch native speakers and the German learners of Dutch L2. An independent t-test showed significant differences between the L1 and L2 ratings on all dimensions. The most pronounced differences were observed in the frequency, familiarity, and usage dimensions, which were assigned much lower values by the L2 learners, and as witnessed by the very large effect sizes (Cohen's  $d > 1.5$ ; Sawilowsky, 2009). In addition, the L2 learners' knowledge was much lower than that of the native speakers. See the Supplementary Materials for the distribution of L1 and L2 intuitions about the idiom properties (Figure S2 and S3 resp.).

We examined the Pearson's correlations between the L1 and L2 intuitions for the different idiom properties (see Table 3.4). For all dimensions, significant correlations were observed. L1 and L2 transparency intuitions showed the strongest correlation (Pearson's  $r = 0.65$ ). High correlations were also observed for imageability and objective idiom knowledge (Pearson's  $r > 0.55$ ), while the intuitions for the dimensions familiarity, frequency, and usage, exhibited much lower correlations (Pearson's  $r < 0.36$ ).

Table 3.3  
Mean and SD for Ratings on Idiom Properties and for Performance on Knowledge Question for Native Speakers and L2 Learners.

<b>Idiom property</b>	<b>Native speakers</b>	<b>L2 learners</b>	<b>t value (df)</b>	<b>d</b>
Frequency	3.77 (0.80)	1.75 (0.61)	12.101 (50.0)	* 2.84
Familiarity	3.43 (0.84)	1.91 (0.78)	7.917 (36.5)	* 1.88
Usage	2.41 (0.87)	1.26 (0.34)	10.589 (50.0)	* 1.75
Transparency	2.54 (0.76)	3.03 (0.80)	-3.322 (36.3)	* 0.64
Imageability	3.29 (0.80)	2.53 (0.73)	3.933 (50.0)	* 0.99
Obj. idiom knowledge	0.88 (0.18)	0.62 (0.26)	10.350 (34.7)	* 1.17

Note. \*  $p < 0.05$

Table 3.4  
Pearson's Correlations between L1 and L2 Ratings per Idiom Property

	<b>L1 - L2</b>
Familiarity	0.195 *
Frequency	0.254 **
Usage	0.356 **
Transparency	0.649 **
Imageability	0.586 **
Objective idiom knowledge	0.563 **

Note. \*  $p < 0.05$  \*\*  $p < 0.01$ .  $n = 110$

### 3.3.3 Intuitions and objective idiom knowledge

To examine to what extent L1 and L2 intuitions of idiom properties reflect L2 idiom knowledge, we carried out logistic mixed effects regression analyses. The responses to the multiple-choice question on idiom knowledge by the L2 learners were converted into a binary variable expressing whether the multiple-choice question was answered correctly or not. This binary variable was used as the dependent variable in the regression analyses.

As explained above, we started off with an initial model only including native predictors: (1) L1 Familiarity, (2) L1 Transparency, (3) L1 Imageability. In addition, we included Participants (random intercept

only), and Idioms (random intercept only) as a random effect. See Table 3.5 for the final model. Because of multicollinearity the predictors L1 Usage and L1 Frequency could not be included in the model.

Transparency as judged by native speakers turned out to be a significant predictor of L2 idiom knowledge. After having established the initial model, we added the same dimensions as rated by the L2 learners. In the presence of these dimensions, the L1 intuitions no longer significantly contributed to the model, and were therefore removed.

The final model included the following predictors: (1) L2 Familiarity, (2) L2 Transparency, (3) L2 Imageability, (4) LexTale score, and (5) Cross-language overlap (reference category: AW). We also added Idioms (random intercept only), and Participants as random effects (random intercept and random slopes of L2 Imageability and L2 Transparency) and standardized the predictors (see Table 3.6).

Because L2 Frequency, L2 Usage and Objective Frequency did not improve the model fit, they were excluded. This was also the case for interactions of intuitions of idiom properties, interactions of idiom properties with LexTale score, and language background variables, such as Number of hours speaking Dutch a week, and Number of years living in the Netherlands.

Positive effects were found for L2 Familiarity ( $\beta = 0.40$ ,  $SE = 0.11$ ,  $p < .001$ ), and L2 Transparency ( $\beta = 2.45$ ,  $SE = 0.20$ ,  $p < .001$ ), while for L2 Imageability a negative effect was found ( $\beta = -0.57$ ,  $SE = 0.15$ ,  $p < .001$ ). We observed a positive effect of vocabulary knowledge as measured by LexTale ( $\beta = 0.61$ ,  $SE = 0.16$ ,  $p < 0.001$ ) and of cross-language overlap. More specifically, if a Dutch idiom was a word-by-word translation of the German expression (AW), L2 learners more often selected the correct meaning in the multiple-choice question than if the Dutch expression had no German equivalent at all (NE) ( $\beta = -0.81$ ,  $SE = 0.39$ ,  $p < .05$ ), the German equivalent consisted of completely different words (DW) ( $\beta = -0.96$ ,  $SE = 0.36$ ,  $p < .01$ ) or if the German equivalent had a number of words in common with the Dutch idiom, but was not a word-by-word translation (nW) ( $\beta = -0.77$ ,  $SE = 0.34$ ,  $p < .05$ ). However, relevelled versions of the model showed that the categories NE, DW, and nW did not significantly differ from each other (NE-DW:  $\beta = -0.16$ ,  $SE = 0.32$ ,  $p = .62$ ; NE-nW:  $\beta = 0.03$ ,  $SE = 0.31$ ,  $p = .92$ ; DW - nW:  $\beta = 0.19$ ,  $SE = 0.28$ ,  $p = .49$ ).

Table 3.5  
Initial Regression Model with L2 Objective Idiom Knowledge as the Dependent Variable  
and L1 Variables as Predictors

<b>Fixed effects</b>	<b>Beta</b>	<b>Std. Error</b>	<b>z value</b>
(Intercept)	0.7052	0.1911	3.690 ***
L1 Familiarity	0.1390	0.1364	1.019
L1 Transparency	0.7581	0.1650	4.595 ***
L1 Imageability	- 0.0668	0.1605	-0.416
<b>Random effects</b>	<b>Variance</b>	<b>Std. Deviation</b>	
Idiom			
Intercept	1.6310	1.2772	
Participant			
Intercept	0.4950	0.7035	

Note: \* $p < 0.05$  \*\* $p < 0.01$  \*\*\* $p < 0.001$ .

Table 3.6  
Final Regression Model with L2 Objective Idiom Knowledge as the Dependent Variable and L2 Variables as Predictors

	<b>Beta</b>	<b>Std. Error</b>	<b>z value</b>
<b>Fixed effects</b>			
(Intercept)	1.9217	0.3383	5.681 ***
L2 Familiarity	0.3976	0.1083	3.672 ***
L2 Transparency	2.4547	0.2034	12.069 ***
L2 Imageability	-0.5693	0.1484	-3.836 ***
LexTale	0.6134	0.1640	3.741 ***
Cross-language overlap NE	-0.8051	0.3862	-2.085 *
Cross-language overlap DW	-0.9637	0.3620	-2.662 **
Cross-language overlap nW	-0.7738	0.3428	-2.257 *
<b>Random effects</b>	<b>Variance</b>	<b>Std. Deviation</b>	<b>Corr</b>
Idiom			
Participant			
Intercept	0.9277	0.9632	
Intercept	0.6155	0.7820	
L2 Transparency	0.7204	0.8488	0.17
L2 Imageability	0.2486	0.4986	0.26
			-0.71

Note. \* $p < 0.05$  \*\* $p < 0.01$  \*\*\* $p < 0.001$ .

### 3.4 Discussion

In the current study we investigated (1) to what extent L2 learners are capable of developing reliable intuitions about idiom properties, (2) how L2 intuitions compare to L1 intuitions about idiom properties, (3) to what extent L2 intuitions better reflect L2 idiom knowledge than L1 intuitions, and (4) to what extent L2 intuitions can provide insights into the mechanisms underlying L2 idiom acquisition. To this end, we collected intuitions about various idiom properties from native speakers and German L2 learners of Dutch for 110 Dutch idiomatic expressions.

#### 3.4.1 Reliability

L2 intuitions about idiom properties turned out to be highly reliable (ICC > .86) and the reliability coefficients were only slightly lower than those obtained for L1 intuitions (ICC > .91). The objective idiom knowledge test also turned out to be reliable for both native speakers and L2 learners (Cronbach's alpha = 0.91).

These findings on the reliability of L1 and L2 intuitions are in contrast with those by Nordmann et al. (2014), who reported low reliability for L1 and L2 intuitions. The L2 learners in their study formed a less homogenous group than those in our study, who all had German as their L1, a language that is relatively close to Dutch. It might be argued that this greater homogeneity led to the high reliability coefficients, but this does not appear to be a plausible explanation, as the native speakers in Nordmann et al. (2014) also constituted a homogeneous group and reliability was low also for that group of participants. A possible explanation for this difference might be the measure used to calculate reliability. Nordmann et al. (2014) used Krippendorff's alpha and interpreted this as a measure of reliability. However, this measure in fact reflects agreement rather than reliability (Tinsley & Brown, 2000). Agreement and reliability measure different aspects of a set of ratings. Agreement has to do with the absolute values of ratings, it indicates to what extent the values are identical. Reliability, on the other hand, indicates to what extent ratings covary (Rietveld & van Hout, 1993; Tinsley & Weiss, 1975). Therefore, low values of Krippendorff's alpha may indicate low agreement, but not necessarily low reliability (see



Chapter 2 and Hubers et al., 2019 for a more elaborate discussion on reliability).

Another important element that probably contributed to the high reliability of our ratings for the native speakers and the L2 learners was the precise and careful way in which the questions about the idiom properties were formulated. Previous studies used varying definitions of the idiom properties under investigation and the questions posed were sometimes ambiguous, leaving room for different interpretations, which may result in more variation in the ratings and lower reliability (see Hubers et al., 2019).

### **3.4.2 Comparison of L1 and L2 intuitions**

L1 intuitions about frequency (3.77), familiarity (3.43), and usage (2.41) exhibited much higher values than those of L2 learners (frequency: 1.75; familiarity: 1.91; usage: 1.26). This is in line with our expectations and with previous studies reporting large differences between L1 and L2 ratings of idiom frequency (Carrol et al., 2017; Nordmann et al., 2014) and familiarity (Nordmann et al., 2014). L2 learners have less experience with the second language than native speakers of that language, and thus with idiomatic expressions (Wray, 2002).

Although a number of studies have examined idiom imageability in native speakers (Cacciari & Glucksberg, 1995), and L2 learners (Steinel et al., 2007), to the best of our knowledge a systematic comparison of L1 and L2 ratings on this dimension has not been conducted before. Based on the limited research on the role of imageability in L2 idiom learning, we expected the L2 learners to rely more on images than the native speakers. However, we found that the latter group rated the idioms as more imageable (3.29) than the L2 learners (2.53) did. Apparently, the native speakers' higher language proficiency and higher familiarity with the meaning of the idioms made it easier for them to visualize the idioms. However, what aspect of the idiom they visualized, the literal or the figurative interpretation, is not entirely clear.

The L2 transparency ratings were higher (3.03) than the L1 transparency ratings (2.54), which is in contrast with studies reporting that native speakers consider idioms to be more transparent than L2 learners do (Abel, 2003; Malt & Eiter, 2004). Keysar and Bly (1995)

argued that transparency intuitions emerge post facto, after participants have learned to associate a given meaning with an idiom, suggesting that transparency intuitions are not necessarily derived from literal meanings. If this is true and transparency intuitions are indeed driven by idiom knowledge, then L1 transparency intuitions should be higher than L2 transparency intuitions. However, our results show the opposite pattern. In our study, we selected the idioms in such a way that transparency, as judged by native speakers in an earlier study, varied. For our question about transparency, we presented the idiom and its meaning and asked people to what extent the individual words could be used to arrive at the figurative meaning. Due to this operationalization people were encouraged to use idiom-inherent properties to rate transparency. Our finding that native speakers did not assign higher transparency ratings to the same idioms than L2 learners therefore lends support to Skoufaki's (2008) more hybrid view of idiom transparency. She suggests that transparency intuitions are also based on more "objective", idiom-inherent properties and are not only developed after participants have learned to associate a specific meaning with an idiom. In addition, the high reliability coefficients obtained for the transparency ratings suggest that L2 learners are capable of consistently rating these intrinsic idiom properties.

The largest differences between L1 and L2 intuitions are observed for the dimensions frequency, familiarity, and usage, whereas for the dimensions transparency and imageability the differences are much smaller. This dichotomy is also visible in terms of correlations: The L1 and L2 intuitions of the dimensions transparency, and imageability are more strongly correlated with each other than the intuitions of the dimensions familiarity, frequency and usage. This suggests that the dimensions frequency, familiarity, and usage are different from transparency and imageability. This difference may lie in the nature of the dimensions. Intuitions of frequency, familiarity, and usage are based on people's experience with the idiom (experience-based dimensions), whereas transparency and imageability intuitions are more related to intrinsic properties of the idioms themselves (content-based dimensions)(see Hubers et al., 2019). It is therefore plausible that the largest differences between the native speakers and L2 learners (in terms

of both the mean ratings and correlations) are observed for the dimensions that are related to language experience.

### **3.4.3 Intuitions and objective idiom knowledge**

Researchers in L2 idiom processing and acquisition often rely on L1 intuitions as a basis for material selection in experiments targeting L2 learners and statistical analyses about L2 idiom processing and learning. We investigated whether L1 intuitions are good predictors of L2 idiom knowledge, or whether L2 intuitions would be preferable. L1 intuitions of familiarity and transparency did affect L2 idiom knowledge in the absence of L2 intuitions. However, after adding the L2 intuitions of the corresponding idiom properties to the analysis, the L1 intuitions lost their predictive power. In other words, L2 intuitions of familiarity, transparency, and imageability seem to be more informative when studying L2 idiom knowledge than L1 intuitions.

The analyses do not only allow us to examine whether L2 or L1 intuitions are better predictors of L2 idiom knowledge, but also give insights into the nature of the relations between L2 idiom knowledge and the intuitions. The final analysis revealed that familiarity, transparency, and imageability (as rated by the L2 learners) affected L2 idiom knowledge. For familiarity and transparency positive effects were observed, while imageability negatively affected L2 idiom knowledge. Transparency most strongly influenced L2 idiom knowledge. L2 learners rely on idiom transparency to arrive at the idiom meaning, because they are less familiar with the expressions than native speakers. The more transparent the idiom, the better the L2 idiom knowledge. Since L2 learners are in general less familiar with the meaning of the idioms, they may be more likely to visualize the literal reading of the idioms. In turn this could hinder them to correctly answer the knowledge question. The negative effect of imageability might seem to contrast with findings indicating that forming an image of the idiom positively affects idiom learning (Steinel et al., 2007). However, as described in the introduction, negative effects of imageability on idiom processing and idiom knowledge have also been observed for native speakers, who have much more experience with idioms (Cacciari & Glucksberg, 1995; Hubers et al., 2019) and L2 learners (Boers et al., 2008). These studies suggest that

people are more inclined to form an image of the literal reading of an idiom.

Vocabulary knowledge (LexTale) and cross-language overlap positively affected L2 idiom knowledge. Although we recruited a homogeneous group of L2 learners in terms of language background and proficiency, vocabulary knowledge was still an important predictor of L2 idiom knowledge. This finding confirms the strong relation between vocabulary knowledge and idiom knowledge, as other studies also pointed out (e.g., Irujo, 1986a; Zyzik, 2011). Zyzik (2011) found that lexical knowledge of single words facilitated idiom learning: Meaning recall for idioms containing unknown words was more difficult than for idioms containing known words.

The effect of cross-language overlap indicates that L2 learners benefit from idioms in their L1 that are word-to-word translations of the idiom in their L2. Surprisingly, we did not find significant differences between the other three categories. Idiomatic expressions that do exist in L1 as an almost, but not exact word-to-word translation did not appear to be better known than idiomatic expressions that do not exist in L1, or do exist in L1, but in totally different words. L2 learners appear to use their L1 idiom knowledge to arrive at the correct idiom meaning in the L2 especially for L2 idioms that have an exact equivalent in the L1. In the situation of the exact equivalents, L2 learners probably feel confident enough to rely on their L1, whereas if the L2 idioms only partially overlap with the L1 equivalents, they are hesitant to resort to their L1 knowledge. These findings complement those of other studies on L2 idioms (Carrol et al., 2017; Charteris-Black, 2002; Irujo, 1986b; Titone et al., 2015) and provide a more fine-grained picture of the impact of cross-language overlap on L2 idiom knowledge.

L2 intuitions of frequency and usage did not significantly affect L2 idiom knowledge. This is in contrast to Hubers et al. (2019), who reported significant effects of L1 frequency and usage on L1 idiom knowledge. The absence of these effects may be due to the relatively low scores and limited variability in the experience-based dimensions familiarity, frequency and usage as rated by the L2 learners, suggesting that familiarity, frequency and usage measure roughly the same construct.

Interestingly, even in the absence of an effect of L2 subjective frequency, we did not find an effect of objective frequency as obtained by corpora. Although an objective measure of idiom frequency positively affects idiom knowledge of native speakers (Hubers et al., 2019), it does not seem a relevant factor in predicting idiom knowledge by L2 learners. Objective idiom frequency may start to positively affect idiom knowledge only after more exposure to the L2. The absence of an effect of objectively measured idiom frequency may be a reason to, in the case of L2 learners, rely more on intuitions obtained from the learners themselves.

#### **3.4.4 L2 idiom acquisition**

The results obtained in the current study can provide insights into the mechanisms underlying L2 idiom acquisition. The lower scores by L2 learners on the experience-based dimensions (frequency, usage and familiarity) could mean that the idioms are just infrequent in the input of L2 learners. Although the same idioms received high scores on these dimensions from native speakers, it is possible that in the input of L2 learners, which is of course more limited than that of native speakers, they are indeed less frequent. Another possible explanation is that the L2 idioms and their figurative meanings are frequent, but that they are not noticed by L2 learners, in other words, the idioms are not so salient. This is in line with previous research showing that L2 learners fail to notice idiomatic expressions in their L2 (Boers & Lindstromberg, 2012; Eyckmans, Boers, & Stengers, 2007; Jones & Haywood, 2004; Peters, 2012). These low scores can be viewed as evidence in favor of a non-formulaic approach to idiomatic expressions in line with Wray (2002, 2008). If the L2 idioms were recognized as chunks, this would result in higher scores on the experience-based dimensions.

Another argument in support of the non-formulaic approach comes from our results with respect to transparency. The correlation between the L1 and L2 transparency ratings was relatively high, indicating that native speakers and L2 learners identify the same idiomatic expressions as transparent or opaque. In addition, the transparency ratings by L2 learners were higher than those by the native speakers, which means that the L2 learners focused more on the individual words in the idioms and on their contribution to the meaning

of the idiom. This is in line with Abel (2003), who found that L2 learners judged non-transparent idioms as more transparent. This, in combination with the finding that transparency was the most important predictor of L2 idiom knowledge, seems to indicate that L2 learners do indeed rely more on the individual words than native speakers, as was suggested by Cieślicka (2006).

However, this does not necessarily mean that L2 learners are not capable of applying a formulaic approach to L2 acquisition. Other results of our study do support a chunking strategy as proposed by Ellis (2001). The fact that L2 learners are capable of developing reliable intuitions about idiom properties related to the idioms themselves (transparency and imageability) suggests that L2 learners manage to identify, and recognize L2 idiomatic expressions, even if these are relatively infrequent and contain words that are not always adjacent and that may differ in form as a result of inflections and conjugations. In other words, it seems that similar acquisition mechanisms are active in L2 learners and native speakers and that the differences observed in knowledge and performance are primarily a question of differences in degree of exposure rather than in mechanisms underlying language acquisition. These findings seem to be in line with those presented by Durrant and Schmitt (2010) with respect to collocation learning.

### **3.5 Conclusions**

In the current study, we investigated to what extent L2 intuitions of idiom properties differ from L1 intuitions in terms of average values and reliability, whether L1 or L2 intuitions are better predictors of L2 idiom knowledge, and to what extent L2 intuitions can provide insight into the mechanism underlying L2 idiom acquisition.

We show that there are differences as the average values of L2 intuitions are significantly lower than those of L1 intuitions. The largest differences were found for experience-based dimensions like frequency, usage and familiarity, but for content-based dimensions like imageability and transparency significant differences were also found. We also found similarities as a homogeneous group of L2 learners appear to be able to develop reliable intuitions about idiom properties, although the reliability values are slightly lower for L2 intuitions than for L1 intuitions.

Importantly, our study also revealed that L2 intuitions are better predictors of L2 idiom knowledge than L1 intuitions.

From a theoretical point of view our study provides important insights into the mechanisms underlying L2 idiom acquisition. L2 learners appear to rely more on the individual words contained in idioms than native speakers (Wray, 2002, 2008), but the high reliability scores of their intuitions suggest that L2 learners manage to identify, recognize and remember L2 idiomatic expressions as chunks, in spite of their relatively variable nature (Ellis, 2001; Ellis & Wulff, 2015). In this respect their strategies do not fundamentally differ from those of native speakers, but only gradually. However, it is clear that further research is needed that more directly taps in L2 idiom processing and acquisition.

From a methodological perspective, these results provide relevant insights that should be taken into account when designing experiments on L2 idiom processing. It is often the case that in such studies the selection of material and the statistical analyses are based on intuitions obtained from L1 speakers. Our study has shown that L1 intuitions are different from L2 intuitions and that the latter are reliable and better reflect L2 knowledge, at least when they are obtained from a relatively homogeneous group of L2 learners with the same L1. Depending on the specific aim of the research, it might be worthwhile to collect and use L2 ratings of idiomatic expressions. In addition, when collecting these ratings, it is important to precisely formulate the rating questions in order to obtain reliable results.

To conclude, the current study shows that L2 intuitions about idiom properties are a valuable and reliable source of information that gives more insight into L2 idiom knowledge and the mechanisms underlying L2 idiom acquisition. Differences between intuitions by L2 learners and native speakers and their relations to idiom knowledge lead us to conclude that for L2 learners the individual words are more salient than the figurative meaning, whereas for native speakers this does not seem the case. However, the figurative meaning may become more salient as a function of exposure. Furthermore, these differences and the finding that L2 learners are able to develop reliable intuitions suggest that L2 intuitions require more attention and should be taken into account when studying L2 idiom processing and acquisition.





## Chapter 4

### **Learning L2 idioms in a CALL environment: The role of practice intensity, modality, and idiom properties**

**This chapter has been reformatted and slightly modified from:**  
Cucchiarini, C., Hubers, F., & Strik, H. (submitted). Learning L2 idioms in  
a CALL environment: the role of practice intensity, modality, and  
idiom properties. *Computer Assisted Language Learning*

### **Abstract**

Idiomatic expressions like *hit the road* or *turn the tables* are known to be problematic for L2 learners, but research indicates that learning L2 idiomatic language is important. Relatively few studies, most of them focusing on English idioms, have investigated how L2 idioms are actually acquired and how this process is affected by important idiom properties like frequency, transparency and cross-language similarity.

The present study employed a specially designed CALL system to investigate the effects of intensity of practice and the reading modality on learning Dutch L2 idioms, as well as the impact of idiom transparency and cross-language similarity. The results show that CALL practice with a focus on meaning and form is effective for learning L2 idioms and that the degree of practice needed depends on the properties of the idioms. L2 learners can achieve or even exceed native-like performance. Practicing reading idioms aloud does not lead to significantly higher performance than reading idioms silently.

These findings have theoretical implications as they show that differences between native speakers and L2 learners are due to differences in exposure, rather than to different underlying acquisition mechanisms. For teaching practice, this study indicates that a properly designed CALL system is an effective and an ecologically sound environment for learning L2 idioms, a generally unattended area in L2 classes, and that teaching priorities should be based on degree of transparency and cross-language similarity of L2 idioms.

### **Acknowledgments**

We are grateful to our fellow members of the ISLA team, Wendy van Ginkel, Ton Dijkstra, Louis ten Bosch, and Alexander Wahl for their valuable input and to Ton Dijkstra for his helpful comments on earlier versions of the manuscript. This work is part of the research program Free Competition in the Humanities with project number 360-70-510 NWO ISLA, which is financed by the Dutch Research Council (NWO).

#### 4.1 Introduction

Formulaic language refers to highly conventionalized ways of expressing meanings through more or less fixed expressions and word combinations that can vary from two-word combinations like collocations (*soft noise*) and binomials (*black and white*) to lexical bundles (*by the way*), phrasal verbs (*put up with*), speech formulas (*I wonder if you'd mind...*), and idioms (*fly off the handle*) (Kecskes, 2007; Wray, 2002). Such formulaic expressions are pervasive in native written language and in spoken discourse (Pawley & Syder, 1983; Siyanova-Chanturia, Conklin, & van Heuven, 2011). They constitute an important part of vocabulary, the “heart and soul of native-like language use”, (Kecskes, 2015, p. 5) and contribute to more fluent language production (Wood, 2006, 2010) and higher L2 proficiency scores (Stengers, Boers, Housen, & Eyckmans, 2010, 2011), while incorrect use of formulaic language negatively impacts oral proficiency (Stengers et al., 2011) and comprehension (Millar, 2010). However, formulaic language appears to be much less frequent in L2 speech and writing, and L2 learners are known to experience difficulties in acquiring formulaic language (Ellis et al., 2008; Kecskes, 2007; Warga, 2005).

A particularly challenging category of formulaic expressions for L2 learners are idiomatic expressions (Cieślicka, 2006; Conklin & Schmitt, 2008; Ellis et al., 2008; Steinel et al., 2007; Weinert, 1995; Wray, 2000), like *to feel under the weather* or *to see eye to eye*. These are generally defined as strings of constituents “whose global meaning does not necessarily derive from that of the constituent parts” (Cacciari, 2014, p. 269). Idioms have received considerable attention in research on L1 comprehension and production, and various models have been proposed to explain L1 idiom processing (e.g., Cacciari and Tabossi, 1988; Sprenger et al., 2006; Swinney and Cutler, 1979; Titone and Libben, 2014). In contrast, L2 idiom processing has been studied less thoroughly.

Because research indicates that a considerable proportion of language is idiomatic in nature (Pawley & Syder, 1983; Siyanova-Chanturia, Conklin, & Schmitt, 2011), that learning idiomatic language has a positive effect on L2 proficiency (Boers, Eyckmans, Kappel, et al., 2006; Hsu & Chiu, 2008) and that L2 learners consider learning L2 idioms

important (Liontas, 2002), it is worthwhile investigating how this process can be facilitated. Pedagogical studies conducted by Boers and colleagues have shown how L2 idiom learning can be stimulated by employing a web-based application (Boers et al., 2004, 2007). However, so far, relatively few studies have investigated how idiomatic expressions are actually acquired in an L2 and how this process is affected by important properties of idioms such as frequency, transparency and cross-language similarity. In addition, the majority of these studies have focused on English idioms (e.g., Ahmadi et al., 2017; Beck and Weber, 2016; Boers et al., 2007; Nordmann et al., 2014; Steinel et al., 2007; Titone and Libben, 2014). However, idiomatic expressions may incorporate specific linguistic and cultural knowledge (Boers et al., 2004; Kovecses & Szabó, 1996) and it is important to gather knowledge about idioms in languages other than English (Bonin et al., 2013; Caillies, 2009; Citron et al., 2016; Nordmann & Jambazova, 2017; Tabossi et al., 2011).

The present study aims at filling in this knowledge gap by addressing the acquisition of Dutch idioms, which have not been studied to such a degree before, and by adopting a twofold approach. First, we investigate the effects of intensity of practice and reading modality on learning idiomatic expressions in Dutch L2 by using a CALL system that provides German L2 learners of Dutch with immediate corrective feedback during practice. To examine the effect of intensity of practice, idioms are divided into two groups: idioms that only receive limited practice, and idioms that receive intensive practice. In order to investigate the effect of reading modality, participants will be instructed to either read aloud the idiom or read it silently after completing the exercises. Second, we study the effect of important idiom properties, such as transparency and cross-language overlap on learning Dutch idiomatic expressions

The advantages of using a CALL system are that learners can work independently at their own pace; intensive practice can be provided; important variables such as practice intensity, modality of practice, and idiom properties can be systematically varied, and that the ecological validity of experimental conditions can be ensured.

#### 4.1.1 Research background

##### *Cross-language overlap*

Early studies on L2 idioms mainly focused on the difficulties L2 learners encounter in comprehending and producing idiomatic expressions in the L2 (Abel, 2003; Irujo, 1986b, 1993; Kellerman, 1979; Pawley & Syder, 1983). One of the important factors affecting L2 idiom comprehension and production is cross-language overlap or similarity. In a study by Irujo (1986), for example, a discourse completion task was used to examine how advanced learners of English with L1 Spanish process English idioms that vary in terms of their cross-language similarity relative to Spanish idioms. In production, it appeared that only identical idioms were easier to process, while similar and different idioms were difficult because of L1 interference. In comprehension, on the other hand, the Spanish learners found L2 idioms that were identical or similar to L1 idioms easier, while idioms that were different were more problematic. Similar results were obtained for fluent Spanish-English bilinguals (Irujo, 1993). However, the professional participants involved in this study showed better performance on L2 idioms that were different from those in their L1. In addition, cross-language similarity was found to be a stronger determinant of L2 idiom production than frequency of use and idiom transparency (i.e., the degree to which the semantic value of the entire expression can be understood in terms of the semantic values of its constituting words; Steinel et al., 2007).

Further evidence attesting the impact of cross-language similarity on L2 idiom production was obtained in a study on idiom avoidance by Laufer (2000) with Hebrew L1 learners of English L2. Avoidance was defined as “a strategy learners may resort to in order to overcome communicative difficulty” (Laufer, 2000, p. 186). Four levels of similarity between L1 and L2 idioms were distinguished: (1) Complete similarity in form (lay the cards on the table), (2) Partial similarity in form (E: miss the boat; H: miss the train), (3) Lack of similarity in form (E: to take someone for a ride; H: to work on someone), and (4) No correspondence (E: it’s not my cup of tea; H: no idiom). The task in this study was elicited production of idioms through a gap-filling translation test. In production, type 2 and type 4 idioms were avoided more than the other two types.

The effect of L1-L2 similarity on L2 idiom processing was further attested by Charteris-Black (2002) who made a clear distinction between conceptual similarity and linguistic similarity. A taxonomy with six different levels was developed to account for (a) correspondence between linguistic form and conceptual basis, and (b) whether idioms are culture specific or universal. Idioms with equivalent concepts and equivalent linguistic forms were the easiest in comprehension (multiple choice task) and production (cued production task), while the most difficult ones were those with different concepts and equivalent linguistic forms and the culture-specific idioms with different conceptual bases and different linguistic forms. On the basis of these results, it was suggested that teaching approaches should prioritize idioms that are different in L1 and L2, and should focus on linguistic forms in the case of conceptual similarity in the two languages.

While various studies underline the impact of cross-language similarity on L2 idiom processing, Türker (2016) has recently shown that the presence of context might mitigate the effect of L1-L2 similarity. In a computer-mediated learning environment, English (L1) learners of Korean (L2) engaged in structured input activities (VanPatten, 1996) that focused on both form and meaning. The environment was especially aimed at establishing form-meaning connections in the absence of explicit instruction. Three types of idioms were investigated: (1) idioms that were identical in form and meaning in L1 and L2, (2) idioms that were identical in form but different in meaning (to have a/the heart; E: to be kind and K: to be willing to do sth.), and (3) idioms that existed in L2 only. The participants were tested on a) idiom production (through gap-filling), b) active recognition (through a multiple choice test), and c) translation into L1 (English). The treatment was found to be effective in several respects. L2 learners managed to learn L2 idioms irrespective of their degree of similarity to L1 idioms if sufficient context was provided. Performance was better for active recognition, followed by translation into L1 and production.

#### *Idiom learning and pedagogy*

Although the majority of early studies on L2 idioms primarily addressed difficulties in L2 idiom comprehension and/or production, there were

also some studies that investigated how idiomatic expressions are learned in the L2.

Boers and Lindstromberg (2012) provide a comprehensive review of pedagogical interventions that have been applied over the years to stimulate the acquisition of formulaic language in an L2. They present an insightful categorization of instructional interventions and their effect on the acquisition of L2 formulaic expressions. Most of the available studies addressed the acquisition of collocations, though some investigations on idiom learning are also reported. What is remarkable about this review and about L2 idiom research in general is that relatively few studies have investigated how actual practice with idiomatic expressions affects their learning.

One such study was conducted by Steinel et al. (2007), who employed a paired-associate learning (PAL) task to teach Dutch (L1) learners of English (L2) 20 English idiomatic expressions varying in Transparency (degree of overlap between the literal and the figurative meaning) and Imageability (the extent to which an idiom can evoke a mental image). The study investigated the effect of direction of learning and testing on L2 idiom comprehension and production, as well as the effect of idiom properties such as transparency and imageability on performance. In the treatment phase, participants saw each idiom and its translation either in a productive order (first L1 idiom, then its L2 translation) or in a receptive order (L2-L1). During testing, they were presented with either the L1 idiom (L1-L2) or the L2 idiom (L2-L1) and were asked to type in its translation. The direction of learning made a difference: Learning idioms productively led to better performance on the productive task, while learning them receptively led to better performance on the receptive task. Performance was strongly affected by Imageability, as indicated by lower scores for low imageable idioms compared to intermediate and high imageable ones. Transparency had a positive effect on comprehension, but not on production.

An interesting series of pedagogical studies were conducted by Boers and colleagues by using a web-based application to provide different types of practice on idiomatic expressions. Boers et al. (2004, 2007) showed that making L2 learners of English aware of the etymology of English idioms favored retention of those expressions. In a subsequent

study by Boers et al. (2008), learners who had exhibited lower performance in a previous study and who had appeared to be low imagers, received additional support in terms of additional pictorial elucidation of the explanation of the idiom's origin (Boers, Eyckmans, and Stengers, 2006). Boers et al. (2008) showed that pictorial elucidation contributed to better retention of the meaning of the idiom, but did not have a positive effect on recollecting the exact linguistic form of the expression. The authors hypothesized that the images might actually distract the learners from paying attention to the linguistic form.

To test this hypothesis, a second learning study was conducted (Boers et al., 2009) that showed that when the linguistic form was presented together with pictures illustrating the idioms, the latter attracted the learners' attention, thus favoring retention of the meaning of the idioms, but not of their linguistic form. In a more recent study, Stengers et al. (2016) modified their web-based learning application in an attempt to engage L2 learners with the linguistic form of L2 idioms through idiom copying. In addition to the online exercises adopted in previous studies, half of the Dutch (L1) learners of English (L2) participating in this study were asked to type each idiomatic expression in a type-in window (an exercise drawing attention to form), while the other half rated the usefulness of each idiom on a 5-point scale (a task that did not focus on form). The copying exercise proved not to be effective in stimulating retention of the linguistic form of the expressions. The authors provide several explanations for these results, including the limited congruence of the copying exercise (which did not require retrieval from memory) with the test (which did appeal to retrieval ability), and the relative shallow form of engagement induced by copy-pasting.

This research overview indicates that several important insights have been obtained on L2 idiom processing and learning, but that many questions remain. For instance, research has shown that practice with idiomatic expressions is beneficial for learning, but it is not clear how intensive this practice needs to be and whether intensity of practice should vary for idioms that vary along different dimensions, such as transparency and imageability. Furthermore, previous studies show that practice should focus on both meaning and form, but do not specify how



this could best be done. The review also shows that cross-language overlap is an important property that affects L2 idioms processing, with idioms that are identical in L1 and L2 being easier to learn than idioms that are different. However, it is less clear how intermediate levels of cross-language similarity affect L2 idiom learning. With respect to cross-language similarity and other idiom properties, such as transparency and frequency, there is still little information on which L2 idioms should be prioritized in teaching and how this should be established.

#### **4.1.2 The present study**

We conducted a CALL-based study examining the effects of intensity of practice and reading modality on L2 idiom learning, as well as the impact of important idiom properties like Transparency and Cross-Language Overlap.

Innovative to this study is that it is couched in a broader project investigating learning, representation, and processing of formulaic language in L1 and L2: the Idiomatic Second Language Acquisition (ISLA) project<sup>1</sup> (Hubers et al., 2018).

Being part of this larger research program results in a number of important benefits. First, we can rely on a large database of normative native data on idiom properties and idiom knowledge (see Hubers et al., 2018, 2019) that has been compiled especially for this project by collecting data on 374 idioms from more than 500 native Dutch (L1) participants and that is being expanded with data from L2 learners of Dutch. The database is used in the present study to select the idiomatic expressions used as learning material and to develop pre-tests and post-tests for our learning experiments. It allows us to select the idioms based on normative data for important idiom properties, such as cross-language overlap, frequency, familiarity, transparency, and imageability. The database also provides us with knowledge about the extent to which each expression is known by native speakers.

Second, a CALL system is part of the ISLA program to which this study is linked. A CALL system offers the possibility that learners can practice independently, at their own pace, and for as long as they want.

---

<sup>1</sup> Project website: <https://isla.ruhosting.nl>

The use of a CALL system in our learning experiment ensures the ecological validity of experimental conditions, in the sense that they could be easily reproduced in teaching practice if the treatment turns out to be successful. Using the CALL system, we can create conditions that favor explicit learning of idiomatic expressions by L2 learners. An important difference between native speakers and L2 learners with respect to idioms is the amount of exposure to and practice with such expressions. Previous investigations indicate that the acquisition of idiomatic language by L2 learners profits from studies abroad offering L2 immersion (Siyanova & Schmitt, 2008; Towell, Hawkins, & Bazergui, 1996) and that implicit learning of lexis and collocations declines with age (Granena & Long, 2013). Against this background, the CALL system allows us to investigate alternative ways of creating conditions that favor explicit learning of idiomatic expressions by L2 learners through more intensive practice than would be possible in traditional teacher-fronted contexts.

This study builds on previous research that employed CALL-based approaches in combination with structured input activities (Türker, 2016) and reinforcement activities with a focus on form (Stengers et al., 2016). As explained by Stengers et al. (2016), the copying activity they applied was not sufficiently effective in drawing attention to form. In this study, we investigate two alternative ways of drawing attention to the linguistic form of the idiomatic expressions: two types of a reading activity, reading aloud and silently. The rationale behind this choice is that reading aloud should involve more engagement with the phonological form of the idiomatic expression, which is likely to leave stronger memory traces and therefore could lead to better retention.

#### **4.1.3 Research questions and hypotheses**

In our study, we addressed the following research questions:

1. To what extent does practice intensity contribute to L2 idiom learning?
2. To what extent is the effect of practice related to idiom properties such as transparency and cross-language overlap?

3. Does reading modality (reading silently or aloud) during practice have a differential effect on L2 idiom learning?

With respect to the first research question, we predict that practice intensity will have a positive effect on learning, with better performance on idioms that are practiced more intensively.

As to the second question, we expect the effect of practice to be modulated by idiom properties such as transparency and cross-language overlap. In particular, we hypothesize that transparent idioms require less intensive practice than opaque idioms. Concerning the effect of cross-language overlap, we expect idioms that do not exist in L1 to be more difficult to learn (Irujo, 1986b). However, based on available research it is difficult to predict how intermediate degrees of cross-language similarity will affect learning. Idioms that exist in L1 should be easier to learn, because L2 learners are at least familiar with the conceptual representation of the idiom (Charteris-Black, 2002). For instance, in learning new words in the L2, conceptual overlap has been shown to have a facilitatory effect (Meade & Dijkstra, 2017). However, learning is also likely to be affected by the degree of similarity in form between the idioms in L1 and L2. L2 idioms that have word-to-word correspondents in L1, like *change hands* in English and Malay, should be easier to learn than L2 idioms that have only a partial overlap in form (E: *iron fist* and M: *nail iron*) or no overlap at all because they contain completely different words (Charteris-Black, 2002). Which of the two latter types is more problematic is difficult to predict. A partial overlap may be expected to have a facilitative role, but it could just as well be confusing to L2 learners.

With respect to the third question, we hypothesize that reading aloud should be more conducive to learning than reading silently for the simple reason that actively pronouncing the words in the idioms requires more phonological processing and possibly attentional capacity than reading them silently.

## **4.2 Method**

### **4.2.1 Participants**

The participants in this study were 42 German L1 learners of Dutch L2 who studied at a University of Applied Sciences in Nijmegen (HAN) and

for whom learning Dutch was a requirement to access university. Their mean age was 21 (SD = 2.1) and their proficiency level was intermediate, B1 according to CEFR. To obtain more objective data on their proficiency level in Dutch, we administered the Dutch version of the LexTale test (Lemhöfer & Broersma, 2012). The average score turned out to be 64.7 (SD = 8.3). This study was ethically assessed and approved by the Ethics Assessment Committee (EAC) of the Faculty of Arts of Radboud University Nijmegen (number 3382).

#### 4.2.2 Material

*Idiomatic expressions.* We selected 60 Dutch idiomatic expressions from our native benchmark database consisting of 374 idioms with scores about idiom knowledge and various idiom properties, such as Familiarity, Transparency and Imageability (Hubers et al., 2019, 2018). We selected expressions that varied in terms of Transparency, and Imageability, and that were generally known by native speakers, as we thought this was an important criterion for teaching L2 idioms: It would be odd to teach L2 learners idiomatic expressions that are not even known by native speakers. See Table 4.1 for the characteristics of the idiomatic expressions. The idiomatic expressions included in the experiment are presented in the Supplementary Materials, Table S4.

Table 4.1

Mean scores and SDs by native speakers for the idiom properties (scale 1-5) and idiom knowledge (% correct) of the idiomatic expressions included in our experiment and for the subsets A and B.

	<b>Complete dataset (n = 60)</b>	<b>Set A (n = 30)</b>	<b>Set B (n = 30)</b>
Frequency	3.43 (0.78)	3.42 (0.80)	3.44 (0.77)
Familiarity	3.17 (0.88)	3.18 (0.65)	3.15 (0.86)
Usage	2.08 (0.72)	2.08 (0.91)	2.07 (0.79)
Transparency	2.87 (0.75)	2.82 (0.74)	2.93 (0.77)
Imageability	3.41 (0.80)	3.23 (0.82)	3.60 (0.75)
Idiom knowledge	79 (0.23)	79 (0.23)	79 (0.23)

For Cross-Language Overlap (CLO), we decided to adopt a four-level classification similar to those applied by Laufer (2000) and Titone et al. (2015): (1) The Dutch idiom does not exist in German (NE), (2) The Dutch idiom does exist in German, but in completely different content words (DW), (3) The Dutch idiom does exist in German and has  $n$  content words in common (nW), (4) The Dutch idiom has a word-to-word correspondent in German (AW). Two German-Dutch bilingual students determined cross-language overlap separately and subsequently compared their scores. Potential differences were resolved by discussing these idioms in more detail.

*Learning exercises.* After consultations with teachers of Dutch L2, we decided to design exercises according to an approach that is widely used in DL2 teaching, the ABCD model by Neuner et al. (1981). Four steps are envisaged in this approach:

- A. Input of new language elements through reading and/or listening
- B. Guided processing through gap-fill or matching exercises
- C. Semi-guided processing through answering questions or finishing sentences
- D. Conversation or discussion, role playing or other more open exercises with a teacher

Steps A-C can be facilitated in a CALL environment, while D can best be practiced with a teacher (Elshoff, 2014). In line with these principles, we developed the following exercises: (1) PAL paradigm (step A), (2) a gap-fill exercise (step B), (3) sentence completion (step C), and (4) idiom selection (step C).

In the PAL paradigm, inspired by Steinel et al. (2007), the idiomatic expressions were presented to the participants along with their meanings in a one-by-one fashion. The idioms were presented at the center of the screen with their meaning directly below. The meanings were provided in Dutch. After 30 seconds the next idiom-meaning pair automatically appeared on the screen. Participants were instructed to carefully read the idioms and their meanings. No explicit task was formulated.

As part of the gap-fill exercise, participants were presented with context sentences containing an idiomatic expression, in which one word

was missing. This word was always a noun that is part of the idiom. Participants were asked to provide the missing word by typing.

For the sentence completion exercise, context sentences were used that ended in an idiomatic expression. Participants were prompted with these context sentences minus the idiomatic expression and were presented with three possible continuations that all contained an idiomatic expressions. They were instructed to type in the appropriate or correct continuation based on the prompt.

In the idiom selection exercise, participants were presented with the meaning of an idiom. They were asked to select the idiomatic expression out of three possible idioms that corresponded to the given meaning by typing. See the Supplementary Materials for screenshots of the various exercises (Figures S4 – S7).

For the gap-fill, sentence completion, and idiom selection exercises, participants received corrective feedback after typing in the answer, indicating both the correct answer and the answer as provided by them (see screenshot in the Supplementary Materials, Figure S8). Subsequently, participants were presented with the context sentence including the correct answer (in case of gap-fill and sentence completion exercises) or with the correct idiom in isolation (in case of the idiom selection exercise). Depending on the condition the idiom was assigned to participants had to read aloud the prompt (as indicated by a record button) or had to read it silently (if record button was absent). A screenshot of both situations is included in the Supplementary Materials (Figure S9 and S10 respectively).

For each idiomatic expression three context sentences were created, in order to prevent participants from associating an idiom to a specific sentence. The context sentences were created in such a way that they biased to the idiomatic interpretation. In addition, the idiom was always presented at the end of the sentence, so that all context sentences were suitable to be used in the sentence completion exercise. The context sentences were used in the gap-fill exercise and the sentence completion exercise.

### 4.2.3 Design and procedure

We adopted a pre-test post-test within subjects design with the variables Intensity of Practice (Intensive and Limited) and Reading Modality (Silent and Aloud). Participants encountered and engaged in the intensively practiced idioms 12 times prior to the post-test, whereas for the idioms that received limited practice this was only 2 times. The effect of Reading Modality was only assessed for idioms that received intensive practice. Reading Modality was manipulated during the gap-fill, sentence completion and idiom selection exercises. Idioms that were read silently and aloud were also encountered 12 times.

The idiomatic expressions were divided into two sets, A and B, of 30 expressions each to be used for the two conditions of intensive and limited practice. These sets were balanced with respect to idiom properties such as frequency, familiarity, usage, transparency, and imageability (see Table 4.1).

In order to investigate Reading Modality, sets A and B, were again divided into two, resulting in sets A1 and A2, and B1 and B2. These subsets contained 15 idiomatic expressions and were all selected in such a way that it was a plausible reflection of complete dataset with respect to idiom properties.

The materials were divided in four master lists in such a way that all materials were counterbalanced. See Table 4.2. We created three versions of each master list, counterbalancing the context sentences over the training sessions. In that way, the same context sentence was not presented in the same exercise in the same training session for all participants. Each participant received one of the 12 lists.

Table 4.2  
Distribution of the materials over de master lists

	<b>Intensive practice</b>		<b>Limited practice</b>
	<b>Silent</b>	<b>Aloud</b>	
Master list 1	B1	B2	A1, A2
Master list 2	B2	B1	A1, A2
Master list 3	A1	A2	B1, B2
Master list 4	A2	A1	B1, B2

The experiment was divided into five weekly sessions: a pre-test, three training sessions using the CALL system, and a post-test. All sessions were located in a computer room where the participants had access to a computer with internet and head set.

*Pre-test.* The pre-test was conducted online through the Qualtrics platform (Qualtrics, Provo, UT). As part of the pre-test, participants filled in a questionnaire and were asked to provide information on their language background. We asked for self-reported proficiency on speaking, reading, writing, and listening (seven-point scale), the number of years they had been learning Dutch, and the number of hours a week they used Dutch outside of class. Subsequently, participants rated each Dutch idiom on five dimensions: Meaning Familiarity, Frequency of Exposure, Frequency of Usage, Transparency, and Imageability. In addition, their idiom knowledge was tested in a multiple-choice question, in which participants had to select the correct idiom meaning out of four alternatives, and in an open-ended question, in which participants were asked to type in the corresponding meaning. For more detailed information about this procedure, see Hubers et al. (2018, 2019). Participants ended the pre-test by filling in the Dutch version of the LexTale, a test to measure vocabulary knowledge (Lemhöfer & Broersma, 2012).

*Training.* The training sessions were conducted online through a CALL system. During the first training session, participants were presented with the PAL paradigm that included both the idioms that received intensive and limited practice. Subsequently, they completed the gap-fill exercise only for the intensive practice idioms. The first training session took about 45 minutes. The second and third training session both consisted of the sentence completion exercise and the idiom selection exercise for the intensive practice idioms only, and took on average one hour.

*Post-test.* The post-test consisted of two parts: a sentence completion exercise without feedback, and a reduced version of the pre-test. Participants started with a sentence completion exercise that included all idiomatic expressions and that was conducted in the CALL environment. Subsequently, their idiom knowledge of all idioms was tested by means of the multiple-choice and open-ended questions also



included in the pre-test through the Qualtrics platform (Qualtrics, Provo, UT). It took the participants on average 75 minutes to complete the post-test.

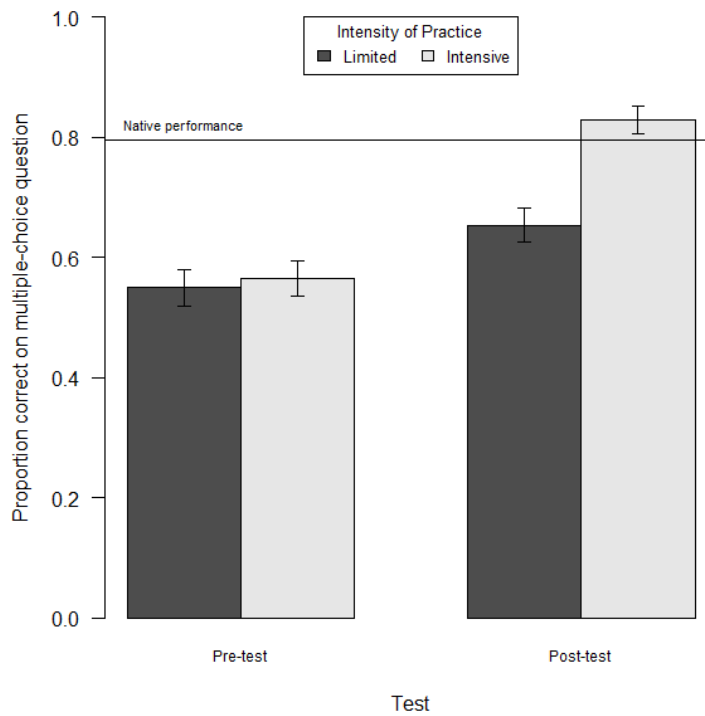
#### **4.2.4 Data analysis**

To examine to what extent Intensity of Practice, Transparency, Cross-Language Overlap and Reading Modality affected idiom learning, we conducted two logistic mixed effects regression analyses. The statistical software package 'R', version 3.4.0 (R Development Core Team, 2008), was used to analyze the data, and the R packages 'lme4' (Bates et al., 2015), 'lmerTest' (Kuznetsova et al., 2017), and 'effects' (Fox, 2003). The regression models were built in a forward manner, starting off with a base model including our variables of interest (Test, Frequency of Practice, Transparency, and Cross-Language Overlap) and a random effect of Participant (random intercept only). The variable Reading Modality was analyzed in a separate regression analysis because this variable is only relevant for a subset of the data (idioms that received intensive practice). The basic model for this analysis included Reading Modality, and Test as fixed effects and Participant as a random effect (random intercept only). Subsequently, we added fixed and random effects one-by-one based on theory and examined whether the model fit improved. If the model fit did not improve, the predictor was not included in the model. The final models are reported in this paper.

### **4.3 Results**

#### **4.3.1 Intensity of Practice, Transparency, and Cross-Language Overlap**

Figure 4.1 shows the mean proportions correct on the multiple choice question targeting idiom meaning recognition split by Test and Intensity of Practice. In the post-test participants performed better than the native speakers in Hubers et al. (2019) on the idiomatic expressions that had received intensive practice.



*Figure 4.1* Mean proportion correct at pre-test and post-test for idioms that received limited and intensive practice. Horizontal line indicates mean native performance for the same idioms as taken from a previous study. The error bars represent SEs.

In order to examine the role of Intensity of Practice, and Transparency and Cross-Language Overlap in idiom learning, we carried out a logistic mixed effects regression analysis. The responses to the multiple-choice question were converted to a binary variable (correct/incorrect) and used as the dependent variable in the analysis.

The fixed effects included in the final model were: (1) Test (reference category: Post-test), (2) Intensity of Practice (reference category: Limited), (3) Transparency, (4) Cross-Language Overlap (reference category: NE), (5) Set (covariate, reference category: A1), (6)

LexTale score (covariate), (7) Test x Intensity of Practice, (8) Test x Cross-Language Overlap, (9) Test x Transparency, and (10) Test x Intensity of Practice x Transparency. Idioms (random intercept only) and Participants (random intercept and random slope of Test) were added to the model as random effects. Adding interaction effects with LexTale and Transparency, and LexTale and Intensity of Practice did not lead to an improvement of the model fit, and were therefore not included in the final model. Although not statistically significant, we included the three-way interaction in our final model, because this resulted in a marginally significant improvement of the model fit ( $\chi^2(2) = 5.444, p = 0.065$ ). The final model is presented in Table 4.3.

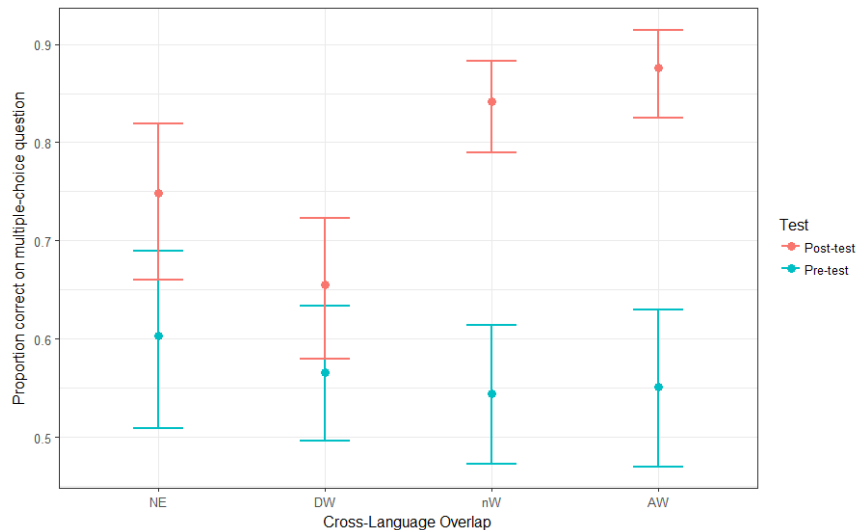
A significant interaction effect of Test and Intensity of Practice was observed ( $\beta = -0.96, SE = 0.14, p < 0.001$ ). In the post-test, participants performed better on idioms that received intensive practice as opposed to idioms that received limited practice ( $\beta = 1.03, SE = 0.11, p < 0.001$ ), whereas in the pre-test, idioms that received limited and intensive practice did not significantly differ (relevelled version of the model:  $\beta = 0.08, SE = 0.09, p > 0.05$ ).

In addition, the analysis revealed a significant interaction effect of Test and Cross-Language Overlap ( $\beta = 0.36, SE = 0.20, p = 0.059$ ). In the pre-test no effects of Cross-Language Overlap were observed, whereas this was the case in the post-test. More specifically, in the post-test, participants performed worse on idiomatic expressions that do not exist in their L1, German, (NE) as opposed to idioms that are word-to-word equivalents (AW,  $\beta = 0.87, SE = 0.30, p < 0.01$ ) and idioms that have a number of content words in common (nW,  $\beta = 0.58, SE = 0.27, p < 0.05$ ). In addition, in the post-test participants more often correctly recognized the meanings of idioms that do not exist in German, than meanings of idioms that do exist in German, but in totally different words (DW,  $\beta = -0.45, SE = 0.22, p < 0.05$ ). A relevelled version of the model showed no differences between idioms that have word-to-word equivalents in German and idioms that have a number of content words in common ( $\beta = -0.29, SE = 0.23, p > 0.05$ ). See Figure 4.2 for a visual presentation of the interaction effect.

Table 4.3 Final Regression Model with Idiom Knowledge as the Dependent Variable

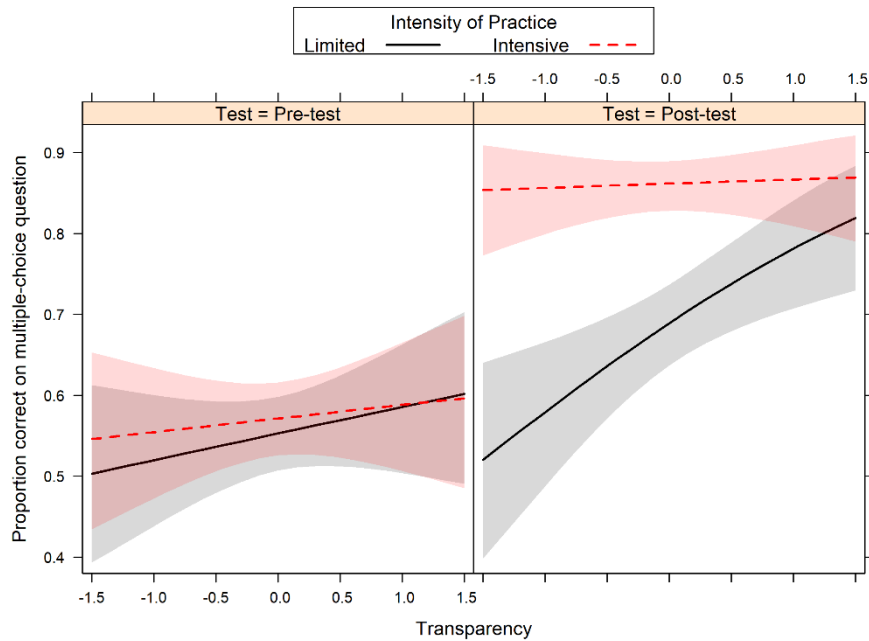
	<b>Beta</b>	<b>Std. Error</b>	<b>z value</b>
<b>Fixed effects</b>			
(Intercept)	0.8577	0.2399	3.575 ***
Pre-test vs. Post-test	-0.1925	0.2057	-0.936
Intensive vs. Limited	1.0321	0.1129	9.143 ***
Transparency	0.4765	0.1529	3.115 **
Cross-language overlap DW vs. NE	-0.4475	0.2202	-2.032 *
Cross-language overlap nW vs. NE	0.5802	0.2681	2.164 *
Cross-language overlap AW vs. NE	0.8703	0.3008	2.893 **
Set A2 vs. A1	-0.4759	0.1745	-2.728 **
Set B1 vs. A1	-0.3903	0.1766	-2.210 *
Set B2 vs. A1	-0.2742	0.1792	-1.530
LexTale score	0.3226	0.0637	5.065 ***
Pre-test x Intensive	-0.9569	0.1437	-6.657 ***
Pre-test x DW	0.2968	0.2067	1.435
Pre-test x nW	-0.8220	0.2502	-3.285 **
Pre-test x AW	-1.0838	0.2809	-3.858 ***
Intensive x Transparency	-0.4328	0.1540	-2.810 **
Pre-test x Transparency	-0.3433	0.1579	-2.174 *
Pre-test x Intensive x Transparency	0.3683	0.1952	1.886 .
<b>Random effects</b>	<b>Variance</b>	<b>Std. Deviation</b>	<b>Corr</b>
Idiom	0.1536	0.3919	
Participant	0.0856	0.2926	
Test	0.1235	0.3514	0.13

Note. \* $p < 0.05$  \*\* $p < 0.01$  \*\*\* $p < 0.001$



*Figure 4.2* The interaction effect Test x Cross-Language Overlap. Effect of Cross-Language Overlap visible for the Post-test only. Error bars represent SEs.

The analysis revealed a marginally significant three-way interaction effect with Test, Intensity of Practice and Transparency ( $\beta = 0.36$ ,  $SE = 0.20$ ,  $p = 0.059$ ) (see Figure 4.3). By looking at the underlying two-way interactions, we gained more insight into the nature of the three-way interaction effect. In the post-test, the effect of Transparency appeared to be significantly smaller for the idioms that were intensively practiced than for the idioms that only received limited practice ( $\beta = -0.43$ ,  $SE = 0.15$ ,  $p < 0.01$ ) (see right panel in Figure 4.3). More specifically, an effect of Transparency was absent for the intensively practiced idioms in the post-test (relevelled version of the model:  $\beta = 0.04$ ,  $SE = 0.17$ ,  $p > 0.05$ ), whereas a significant positive effect was observed for idioms that received limited practice ( $\beta = 0.48$ ,  $SE = 0.15$ ,  $p < 0.01$ ). However, a relevelled version of the model revealed no such differences in the pre-test ( $\beta = -0.06$ ,  $SE = 0.12$ ,  $p > 0.05$ ). Here, an effect of Transparency was absent for both limited ( $\beta = 0.13$ ,  $SE = 0.14$ ,  $p > 0.05$ ) and intensively practiced idioms ( $\beta = 0.07$ ,  $SE = 0.14$ ,  $p > 0.05$ ).



*Figure 4.3* The interaction effect Test x Intensity of Practice x Transparency. Positive effect of Transparency visible only in the post-test (right panel) for the idioms that received limited practice. Error bands are based on SEs.

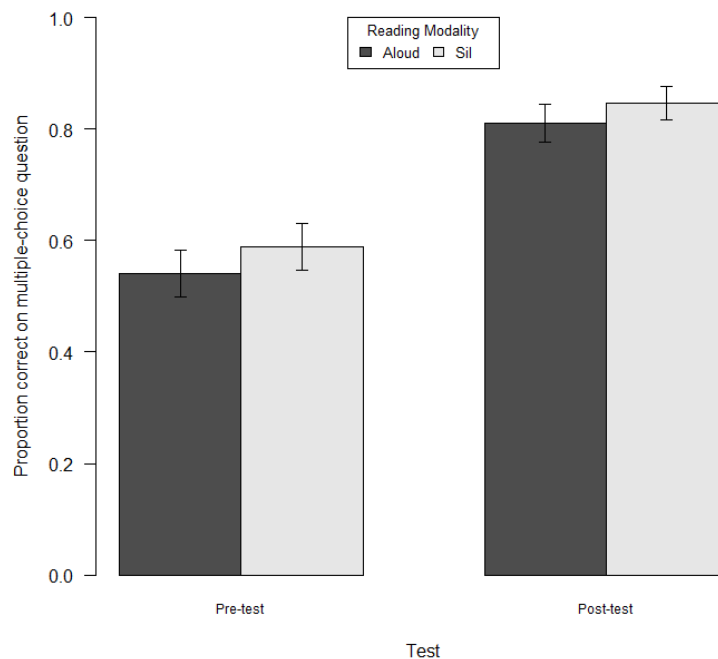
#### 4.3.2 Reading Modality

Figure 4.4 presents the mean proportions correct on the multiple choice question split by Test and Reading Modality only for a subset of the data (i.e. the idiomatic expressions that received intensive practice).

To investigate whether Reading Modality, either silent or aloud, has a differential effect on L2 idiom learning, the pre-test and post-test scores were submitted to logistic mixed effects regression analysis. Again, the performance on the multiple-choice question (correct/incorrect) was used as a dependent variable in the analysis.

The final model included the predictors (1) Test (reference category: Post-test), (2) Reading Modality (reference category: Silently), (3) LexTale score (covariate), and (4) Test x Reading Modality as fixed effects. Participants (random intercept and random slope of Test) and

Idioms (random intercept only) were included as random effects (See Table 4.4). We did not include Transparency and Cross-Language Overlap, because we did not expect these variables to affect reading aloud and reading silently.



*Figure 4.4* Mean proportion correct at pre-test and post-test for intensive practice idioms that were read aloud and silently. The error bars represent SEs.

The analysis revealed only significant effects for Test and LexTale score. Participants performed worse on the pre-test than on the post-test for both the idioms that were read aloud ( $\beta = -1.56$ ,  $SE = 0.19$ ,  $p < 0.001$ ) and read silently ( $\beta = -1.62$ ,  $SE = 0.19$ ,  $p < 0.001$ ). Vocabulary knowledge, as measured by LexTale, positively influenced performance on the multiple-choice question ( $\beta = 0.04$ ,  $SE = 0.009$ ,  $p < 0.001$ ). No significant interaction effect was observed between Reading Modality and Test ( $\beta = 0.06$ ,  $SE = 0.21$ ,  $p > 0.05$ ).

Table 4.4  
*Final Regression Model with Idiom Knowledge as the Dependent Variable*

<b>Fixed effects</b>	<b>Beta</b>	<b>Std. Error</b>	<b>z value</b>
(Intercept)	-0.3609	0.6350	-0.568
Pre-test vs. Post-test	-1.6185	0.1938	-8.353 ***
Aloud vs. Silently	-0.2809	0.1687	-1.665
LexTale score	0.0369	0.0094	3.933 ***
Pre-test vs. Post-test x Aloud vs. Silently	0.0575	0.2108	0.273
<b>Random effects</b>	<b>Variance</b>	<b>Std. Deviation</b>	<b>Corr</b>
Idiom			
Intercept	0.2018	0.4492	
Participant			
Intercept	0.0985	0.3138	
Test	0.4202	0.6482	0.33

Note. \* $p < 0.05$  \*\* $p < 0.01$  \*\*\* $p < 0.001$



#### 4.4 Discussion

In this paper, we employed a CALL system that provides automatic corrective feedback to investigate the effect of intensity of practice and reading modality on learning Dutch L2 idiomatic expressions, as well as the impact of relevant idiom properties like degree of transparency and cross-language overlap.

The intensive practice focused on meaning and form that was provided through our CALL system turns out to impact L2 idiom learning, and L2 learners could even achieve native-like performance. Even limited practice involving only two presentations of each idiomatic expression turned out to be effective. However, the degree of improvement appears to vary for different types of idioms. In particular, while intensive practice was effective for both opaque and transparent idioms, limited practice was effective for the more transparent idioms, but not for the more opaque ones. These findings are in line with our research hypotheses about the positive effect of practice and its relation to idiom transparency.

With respect to the effect of cross-language similarity, we found that L2 learners had more difficulties with idioms that do not exist in their L1, than with idioms that have word-to-word correspondents and idioms that have a number of content words in common. These findings align with our research hypothesis. However, the L2 learners more easily acquired idioms that do not exist in their L1, (Dutch: 'goed uit de verf komen' = to make a good impression) than idioms that do exist in German, but in totally different words (Dutch: 'water naar de zee dragen' = German: 'Eulen nach Athen tragen'; English equivalent: 'to carry coals to Newcastle'). On the one hand, this seems to contrast with findings by Charteris-Black (2002) that idioms with equivalent concepts were easier to learn. On the other hand, it is in line with results by Irujo (1986), who found that only idioms that were identical or similar in form were easier to comprehend, whereas idioms that were different in form were more problematic. The present study adds to this line of research by indicating that idioms with no overlap in form are also more difficult to learn and require more intensive practice.

A somewhat surprising result is that both transparency and cross-language overlap did not have a significant impact on idiom recognition at pre-test. In the post-test we found a significant effect of transparency in the limited practice condition, but not for intensive practice. Thus, it appears that the participants in this study did not exploit this knowledge in the first place, but actually learned to make use of this information as a result of practice with the idiomatic expressions. In addition, the effect of transparency is visible in the limited practice condition where L2 learners have probably become aware of this property and need to use it to arrive at the meaning of L2 idioms. In the intensive practice condition they apparently received enough practice to learn the idioms without having to resort to transparency. Similar results with respect to cross-language overlap were obtained by Türker (2016), who also found no significant differences at pre-test between three types of idioms that were either (a) identical in form and meaning, (b) different in meaning, but identical in form or (c) existed in L2 only.

As to the manipulation of reading modality, no significant differences were found between the conditions in which L2 learners practiced with reading aloud or reading silently. As explained above, we expected reading aloud to have a more positive effect on L2 idiom learning, but this was not borne out by the results. We also expected this type of reinforcement activity to be more effective than the copying task investigated by Stengers et al. (2016), exactly for the reasons these authors mentioned. With reference to Levels of Processing Theory (Cermak & Craik, 1979; Craik & Lockhart, 1972), they argued that the copy-pasting activity was probably not engaging enough and therefore unlikely to leave persistent memory traces (Stengers et al., 2016, p. 296). As a possible, more elaborative alternative, they suggested copying by typing or by handwriting, as these would require more focus on the linguistic form than simple copy-pasting as applied in their study. Both our reading activities, silently and aloud, are tasks that draw attention to the orthographic form, with reading aloud requiring more phonological processing than silent reading. It is therefore possible that in this respect the two tasks were not sufficiently different from each other to bring about a difference in performance. Another possible explanation, also suggested by Stengers et al. (2016) with respect to their findings, is that

the idiom recognition test in the post-test was not completely appropriate to measure the effect of this kind of reinforcement practice. In order to check whether an effect of reading modality was observed in another type of exercise in the post-test, the sentence completion exercise, we examined these results in a post-hoc analysis. The same pattern of results was found for both reading modalities. For this type of exercise, proportion correct for the idioms that had been practiced by reading aloud was 0.74, while the proportion correct was 0.77 for idioms that had been read silently. Finally, another possible explanation for the lack of a facilitatory effect of oral reading practice could be that the amount of oral practice was not sufficient within the context of our learning experiment.

Our results indicate that with sufficient exposure and practice, L2 learners demonstrate a level of meaning recognition equivalent to that of native speakers. This suggests that the architecture underlying idiom comprehension in the two groups may be similar. In fact, recent hybrid models of idiom processing in both L1 and L2 highlight the role of idiom properties, such as Transparency and Cross-Language Overlap, during idiom processing (Libben & Titone, 2008; Titone et al., 2015). These factors appear to exert a similar influence on idiom processing and on L2 idiom learning as observed in our study: Transparency is found to facilitate idiom processing in both native speakers (e.g. Libben and Titone, 2008; van Ginkel and Dijkstra, 2019) and L2 learners (Titone et al., 2015), and Cross-Language Overlap positively affected L2 idiom processing (Titone et al., 2015). Therefore, differences between the two groups seem to result from differences in exposure, rather than from differences in the underlying mechanisms involved in idiom processing.

Additional research is required to shed more light on the lack of a facilitatory effect of oral reading practice. In our study, the specific implementation of the reinforcement activity - reading aloud or silently - may not have been optimal or too short in duration to show any effect. As the current study only included comprehension activities in the pre- and post-test, it may be worthwhile for future research to include idiom production activities as well, as these would give a more complete picture of L2 idiom learning. Another factor that could be included in future studies is linguistic distance. The L1 and L2 involved in this study are

rather close, which is known to facilitate learning. It would be interesting to conduct similar experiments with an L1 and L2 that lie further apart, to see how transparency and cross-language overlap affect idiom learning in that case. Finally, in our future research we intend to investigate idiom learning after incorporating automatic speech recognition in the CALL system, which allows a more natural and possibly more rewarding form of interaction for the learner.

#### **4.5 Conclusions**

The present study allows us to conclude that CALL practice with a focus on meaning and form is effective for learning L2 idioms and that the degree of practice needed depends on the properties of the idioms. L2 learners can achieve or even exceed native-like performance. Higher intensity of practice leads to significantly better performance. Limited practice is effective for learning idioms that are transparent and idioms that contain similar words in L1. With intensive practice the differences in gains are smaller. Put otherwise, more practice is required for opaque idioms, for idioms that do not exist in L1, and for those that do exist in L1, but with a completely different form. Practicing reading idioms aloud does not lead to significantly higher performance than reading idioms silently.

The study has implications from both a theoretical and a teaching perspective. From a theoretical perspective, the findings indicate that differences between native speakers and L2 learners result from differences in exposure, rather than from differences in the underlying mechanisms involved in idiom acquisition. With respect to teaching practice, the study provides insights into which properties make L2 idioms more difficult to learn and how this knowledge could be exploited in idiom teaching. In teaching, specific attention should be paid to opaque idiomatic expressions, to idioms that do not exist in the learner's L1, and to idioms that do exist in the L1, but with a completely different form. Last but not least, the study presents an effective and ecologically valid way of facilitating L2 idiom practice, which is usually an unattended area in L2 classes, through a properly designed CALL system. L2 learners can learn from only two idiom presentations and with intensive practice they can reach native-like proficiency.

## Chapter 5

### **Idiom processing by native speakers: Individual word activation and the role of word frequency**

**This chapter has been reformatted and slightly modified from:**  
Hubers, F., Cucchiaroni, C., Strik, H., & Dijkstra, T. (In preparation).  
Individual word activation and word frequency effects during the  
processing of opaque idiomatic expressions

## **Abstract**

Many studies investigated the relation between the idiomatic expression as a whole and its component words. Recent studies suggest that literal and figurative processing run in parallel. Depending on the time-course and properties of the idiom, the meanings of the individual words are activated or not (Beck & Weber, 2016a; Libben & Titone, 2008; Sprenger et al., 2006; Titone & Libben, 2014). During the processing of opaque idiomatic expressions, for example, the individual word meanings are found to be suppressed (Rommers et al., 2013).

The present study investigated the activation of the individual words during the processing of opaque idiomatic expressions using a task assessing both form and meaning activation. In two primed word-naming experiments, inspired by the paradigm used by Rommers et al. (2013), Dutch native speakers were presented with sentences in a word-by-word fashion and were instructed to read aloud the last word of the sentence as quickly as possible. The target words were either embedded in an idiomatic context or a literal context.

Our results show that the individual words are in competition with the idiom as whole at both the semantic and the orthographic or form level of representation. At the meaning level, activation of the idiom-final noun was not activated, while at the orthographic level, idiom-final noun activation was observed as reflected by an inhibitory effect of word frequency.

These results provide evidence in favor of a hybrid model of idiom processing in which the individual words and the idiom as a whole are in interaction on both the orthographic and semantic level of representation.

## **Acknowledgments**

We are grateful to our fellow members of the ISLA team, Wendy van Ginkel, Louis ten Bosch, and Alexander Wahl for their valuable input. This work is part of the research program Free Competition in the Humanities with project number 360-70-510 NWO ISLA, which is financed by the Dutch Research Council (NWO).

## 5.1 Introduction

There is a longstanding tradition of research on idiom processing in psycholinguistics. One of the main questions in this domain is whether idiomatic expressions, such as 'kick the bucket' and 'spill the beans', are stored as a whole in the mental lexicon or not. A considerable body of evidence has demonstrated that, to some extent, this indeed is the case (e.g., Bobrow & Bell, 1973; Cutting & Bock, 1997; Rommers, Dijkstra, & Bastiaansen, 2013; Sprenger, Levelt, & Kempen, 2006; Swinney & Cutler, 1979; van Ginkel & Dijkstra, 2019).

However, even if idiomatic expressions are stored as a whole, they are still composed of parts: namely, their individual words. This leads to the interesting question of how processing is affected by the relation between those parts (words) and the idiomatic expression as a whole. Take, for instance, an idiomatic expression that can also be literally interpreted: 'to kick the bucket'. In its literal interpretation, the meaning of the target word 'bucket' must be integrated within the literal meaning of the phrase as a whole. How quickly and smoothly this can be done will co-depend on lexical properties of the word 'bucket', for instance, its word frequency. A higher target frequency should lead to *faster* lexical-semantic integration. However, to understand the idiom 'to kick the bucket' in its sense 'to die', the meaning of the individual word 'bucket' is actually interfering with that of the idiom as a whole. Nevertheless, to verify that the idiom 'to kick the bucket' is being presented, the word form of 'bucket' must still be identified. Therefore, when interpreting the expression 'to kick the bucket' as an idiom, it would be best to process the word 'bucket' at an orthographic form level, but to avoid deriving its meaning – if that is possible.

In any case, the processing difficulty of the idiom as a whole should co-depend on the properties of the target word related to the orthography, such as its frequency relative to the frequency of the idiomatic expression as a whole. In the idiomatic case a higher target word frequency should actually lead to *slower* processing of the idiom, which suggests that the idiom as a whole and the target word are in competition at the orthographic level of representation. The role of individual word frequency during idiom processing has received only

little attention (Cronk, Lima, & Schweigert, 1993; Libben & Titone, 2008; van Ginkel & Dijkstra, 2019).

In sum, the idiom as a whole and its individual words are in competition at both the semantic and the orthographic level of representation. Previous studies investigating the role of the individual words during idiom processing have mainly focused on the activation of their semantics (e.g., Cutting & Bock, 1997; Rommers et al., 2013; Sprenger et al., 2006). In addition, semantic idiom properties, such as transparency (i.e., the extent to which the individual word meanings contribute to the figurative meaning; Cacciari, 2014) and familiarity (i.e., the extent to which people are familiar with the idiom's meaning; Hubers, Cucchiari, Strik, & Dijkstra, 2019; Nordmann, Cleland, & Bull, 2014) have been found to modulate the extent to which the individual word meanings are activated (e.g., Titone & Libben, 2014).

In contrast to previous research, the current study aims to shed more light on the interaction between the individual words and the idiomatic expression as a whole at both the orthographic and the semantic level in idiom processing. Before we introduce the present study, we discuss previous studies on the activation of individual words during idiom processing. First, we focus on research that addressed individual word activation at the semantic level and the way this is affected by properties related to the semantics of the idiom as a whole. Next, we review the limited number of studies that examined the activation of individual words during idiom processing at the orthographic level by looking at word frequency effects.

### **5.1.1 Semantic effects of individual words during idiom processing**

Previous research has extensively investigated the relation between the idiomatic expression as a whole and its component words, focusing especially on the semantic activation of the individual words.

Early studies argued that individual word meanings are not activated during idiom processing and that idioms are stored as a whole in the mental lexicon (Bobrow & Bell, 1973; Gibbs, 1980; Swinney & Cutler, 1979). One such non-compositional view was put forward by Bobrow and Bell (1973). According to their Idiom List Hypothesis, idiom comprehension requires a special idiom mode of processing. Once



participants are in this idiom processing mode, the individual word meanings do not affect processing. The Direct Access Hypothesis, proposed by Gibbs (1980), does not identify different processing modes, but suggests that an idiom's figurative meaning can be directly accessed in the mental lexicon without an analysis of the literal meaning. Only if idiomatic processing fails, will phrases be analyzed literally.

Later studies, however, have shown that the semantics of the individual words in idiomatic expressions do contribute to their figurative interpretation. This led to the development of compositional models of idiom processing (Cacciari & Glucksberg, 1991; Cacciari & Tabossi, 1988; Gibbs, Nayak, Bolton, & Keppel, 1989; Nunberg, 1979). Here individual words are activated during idiom processing and an idiom's figurative meaning is retrieved by combining the semantics of the individual words. Prominent compositional models are the Configuration Hypothesis (Cacciari & Tabossi, 1988) and the Idiom Decomposition Hypothesis (Gibbs, Nayak, & Cutting, 1989). The Configuration Hypothesis suggested that idioms do not have a separate entry in the mental lexicon. Individual words are accessed during the processing of both the idiomatic and the literal meaning. In addition, idiomatic meaning becomes available only after information is processed that identifies the phrase as an idiom. Such information is referred to as the idiom's 'key'. According to the Idiom Decomposition Hypothesis, individual word meanings can be used especially if the idiom is analyzable or decomposable (Gibbs, Nayak, & Cutting, 1989). If the idiom is non-decomposable, however, one will try to use the semantics of the individual words, but, in the end, resort to a holistic analysis.

More recent studies suggest that literal and figurative processing run in parallel and, depending on the time-course and properties of the idiom, the meanings of the individual words are activated or not (Beck & Weber, 2016a; Cutting & Bock, 1997; Libben & Titone, 2008; Sprenger et al., 2006; Titone et al., 2015; Titone & Libben, 2014). Sprenger et al. (2006) proposed a hybrid model of idiom production in which idiomatic expressions have separate representations (superlemmas) that are connected to simple word lemmas on the one hand and to idiomatic meaning representations on the other hand. The superlemmas can be accessed by activating the simple lemmas of the component words.

The extent to which the individual word meanings are activated may be modulated by properties related to the semantics of the idiom as a whole. This idea has been put forward by Libben and Titone (2008) in their Constraint-Based Model. Idiom properties such as familiarity and predictability, which are related to direct retrieval, may affect early stages of idiom comprehension, whereas decomposability or transparency may become important later on (Titone et al., 2015; Titone & Libben, 2014). Evidence in support of this model has been provided by Titone and Libben (2014) in a cross-modal priming experiment. They presented participants with auditory idiom-bearing sentences (*They were two peas in a pod*) and control sentences (*It was a very nice chair*). Participants were instructed to perform a lexical decision on a target word semantically related to the idiomatic meaning (*similar*) that was displayed at the offset of the penultimate word, at the offset of the prime sentence, or 1000 ms post offset of the prime sentence. Titone and Libben (2014) observed a significant idiom priming effect: Responses to the target word were faster after auditory presentation of the idiom-bearing sentences than after control sentences. This effect became larger over time. In addition, Titone and Libben (2014) observed a modulation of the idiom priming effect by idiom familiarity, semantic decomposability, and literal plausibility at different points in time. Literal plausibility negatively affected idiom priming at the penultimate position of the idiom, while familiarity positively affected idiom priming at the offset of the idiom, and a negative effect of semantic decomposability was observed 1000 ms post offset of the idiom. Thus, in their study, Titone and Libben (2014) showed that multiple idiom properties jointly affect idiom meaning activation.

To summarize, the semantic activation of individual words during idiom processing has received much attention and has led to three types of idiom processing models. Early studies did not directly assess the activation of individual word meanings, but rather focused on the processing of idiomatic and literal phrases as a whole (e.g., Gibbs, Nayak, & Cutting, 1989; Swinney & Cutler, 1979). Later studies *did* examine the semantic activation of individual words using priming paradigms by assessing semantic spreading activation: If the semantics of a single word (that is part of the idiom) is activated, it should co-activate words that are

semantically related to this word (Cacciari & Tabossi, 1988; Rommers et al., 2013; van Ginkel & Dijkstra, 2019). Finding a facilitatory effect for words that are semantically related to the individual component words (the literal meaning) implies that those component words are semantically activated themselves. Absence of this spreading activation effect during idiom processing has been taken as evidence for suppression of the individual words. Any facilitatory effects have been reduced because the individual words are suppressed.

A study based on this argumentation is the combined RT and EEG study by Rommers et al. (2013). They investigated literal word meaning activation during the processing of Dutch opaque idioms. Participants were presented with idiomatic and literal sentence contexts in a RSVP fashion. The idiomatic sentence contexts always included an idiom (e.g., *After many transactions the careless scammer eventually walked against the lamp yesterday*). Following Federmeier and Kutas (1999), the critical word was either a correct and expected word (COR; *lamp*), a word that was semantically related to the expected word (REL; *candle*), or a word that was semantically unrelated to the expected word (UNREL; *fish*). In the idiomatic sentence contexts, the correct and expected word was always a noun that was part of the idiom. The same critical words were used in literal sentence contexts in which the correct and expected word was equally predictable (e.g., *After lunch the electrician screwed the new light bulb into the lamp yesterday*). In the behavioral version of the experiment, participants performed a lexical decision task on the critical words, while in the EEG version of the experiment, no task was involved and N400 effects were measured. In the literal sentence contexts a graded pattern was observed in terms of response times and N400 effects: The correct and expected word (COR) was responded to the fastest and elicited the smallest N400 effect followed by the semantically related (REL) and unrelated (UNREL) word respectively. In the idiomatic sentence context, however, no difference was observed between the REL and UNREL conditions. Apparently, in the idiomatic sentence context, spreading activation from the expected to the semantically related word was absent. Rommers et al. concluded that “when reading predictable and opaque idiomatic expressions, for which literal word meanings are

irrelevant, the processing of literal word meanings can to some extent be ‘switched off’ ” (2013, p. 775).

### 5.1.2 Orthographic effects of the individual words during idiom processing

If individual words are not accessed during idiom processing, effects of orthographic properties of these words, such as word frequency, should be absent too. If the individual words *are* activated, but activation is not strong enough to access their semantics and subsequently co-activate semantically related words, orthographic effects, such as word frequency, might nevertheless be visible. However, the activation of the orthography of individual words in idiom processing has received limited attention. Only three studies have examined this issue by investigating the role of individual word frequency in idiom processing (Cronk et al., 1993; Libben & Titone, 2008; van Ginkel & Dijkstra, 2019).

Cronk et al. (1993) investigated the role of word frequency in relation to idiom familiarity in a self-paced reading paradigm. Idiom familiarity was obtained through a norming study, in which participants rated how often the phrase (the idiom) is heard used figuratively on a 5-point scale. Frequencies of the idioms’ component words were taken from Kučera and Francis (1967) and were averaged per idiom. Cronk et al. (1993) found that high-familiar idioms were read more quickly than low-familiar idioms, and that this effect was modulated by word frequency: The familiarity effect was larger for idioms containing high frequency words as opposed to idioms containing low frequency words. More specifically, mean reading times per word were much faster for highly familiar idioms consisting of high-frequency words than for high-familiar idioms containing low-frequency words and low-familiar idioms.<sup>2</sup> These findings suggest that the individual words do contribute to the figurative meaning. If the idiom component words are highly

---

<sup>2</sup> We noticed a discrepancy between the results as presented in the original paper by Cronk et al. (1993) and the discussion about these results provided by Libben and Titone (2008). We based our discussion on the results as reported in the original paper.

frequent, the figurative meaning may be retrieved faster as compared to if the component words are of low frequency.

In a series of three experiments, Libben and Titone (2008) investigated the role of various idiom properties, such as familiarity, decomposability, and literality on idiom processing, and the effects of verb and noun frequency in idioms with a 'she [verb] x [noun]' structure. No effect of noun frequency on the RTs for idioms was found. Verb frequency, however, turned out to negatively affect idiom processing, indicating that, paradoxically, idioms with low-frequency verbs were recognized more quickly than idioms with high-frequency verbs. Based on their results, Libben and Titone (2008) argued that infrequent verbs are probably more predictive of idiomatic completions than high-frequency verbs and therefore lead to a processing advantage.

In a priming study, van Ginkel and Dijkstra (2019) presented participants with idiomatic expressions as primes after which target words followed that were either figuratively related to the idiom as a whole (FIG condition), semantically related to the literal word at the end of the idiom (LIT condition), or unrelated to the idiom and the idiom-final noun (UNREL condition). Participants were instructed to perform a lexical decision on the target words. Van Ginkel and Dijkstra (2019) found an idiom priming effect in the FIG condition as compared to the UNREL condition, which they interpreted as evidence in support of the hypothesis that the representations of idioms are activated as a whole. However, they also found that literal word meanings were activated, as reflected by a priming effect for the LIT condition as compared to the UNREL condition. Interestingly, in the LIT condition, a word frequency effect of the idiom-final noun was observed that was absent in the FIG and UNREL condition. More specifically, idiom-final noun frequency negatively affected reaction times on target words semantically related to the idiom-final noun: Higher frequencies resulted in slower reaction times. Van Ginkel and Dijkstra (2019) suggested that this inhibition effect may be due to conflicting processes. On the one hand, the idiomatic reading leads to strong activation of the idiom representation as a whole, while on the other hand, the literal words also become activated. If the idiom-final word is of high frequency, it is more difficult to suppress its

activation than when it is of low frequency. Thus, literal words are not fully suppressed.

Although the studies reviewed above found that individual words are activated during idiom processing, at least at the orthographic level of representation, they showed mixed results with respect to the role of individual word frequency. Cronk et al. (1993) reported a facilitatory effect of word frequency. In contrast, Libben and Titone (2008) found an inhibitory effect of verb frequency on idiom processing, but no effect of idiom-final noun frequency, while van Ginkel and Dijkstra (2019) observed an inhibitory effect of idiom-final noun frequency on idiom processing. These inconsistent results may be due to the different tasks used in these studies. In line with this observation, Van Ginkel and Dijkstra (2019) proposed a context-sensitive hybrid task-dependent processing account, in which literal and figurative processing run in parallel. In this account, the crucial element is the moment at which the target word is presented in relation to the activation of the sentence as a whole.

### **5.1.3 The present study**

To gain more insight into the mixed results of earlier studies, we investigated the role of the individual words during idiom processing at the semantic and orthographic level of representation. With respect to our design, we were inspired by the study of Rommers et al. (2013), who used an RSVP paradigm to investigate the activation of the idiom-final nouns of opaque idiomatic expressions in highly biasing contexts. They found that the activation of the idiom-final nouns was suppressed in terms of their semantics. However, the individual idiom-final words are expected to be activated to some extent, because the word form needs to be identified in order to complete the idiom. Although Rommers et al. (2013) observed no activation of the semantics of the individual idiom-final words during the processing of opaque idiomatic expressions in highly biasing context sentences, effects of lexical properties of the idiom-final nouns related to the orthography, such as word frequency, may still be present.

In order to investigate the activation of the individual words at the semantic and the orthographic level during the processing of opaque idioms, we used the same paradigm as Rommers et al. (2013). However,

instead of a lexical decision task, which also taps into semantic information, we used a word naming task, which relies more on orthography. By focusing more on the word form, effects of the individual word semantics are expected to be reduced, whereas word frequency information, related to the word form, may be retrieved anyway.

Participants were presented with target words embedded in an idiomatic context sentence (“de getrainde dief liep uiteindelijk toch tegen de ...”), or a literal context sentence (“het kind kan niet slapen zonder licht van een kleine ...”). These target words were either the correct and expected target words given the context (COR; “lamp”), semantically related to the expected target word (REL; “warmte”), or semantically unrelated to the expected target word (UNREL; “helm”). The expected target word (COR) in the idiomatic context was always a noun that was part of an idiom (“tegen de lamp lopen”), while the literal context sentences contained a bias to the literal meaning of this same target word.

In line with others using this RSVP paradigm in combination with highly biasing context sentences (Federmeier & Kutas, 1999; Federmeier, McLennan, Ochoa, & Kutas, 2002; Rommers et al., 2013), we hypothesized that participants would respond fastest to the correct and expected target word (COR) in both the idiomatic context (idiom-final noun), and the literal context, as opposed to the semantically unrelated target word and the target word semantically related to the literal meaning of the expected target word. In addition, replicating Rommers et al. (2013), we expected the semantically related target word (REL) and the unrelated target (UNREL) not to differ in the idiomatic context. The activation of the idiom-final noun was expected not to be strong enough to activate its semantics and consequently to co-activate words that are semantically related to the literal interpretation of idiom-final noun, because we included opaque idiomatic expressions, in which the individual word meanings do not contribute to the figurative meaning, and presented them in a highly idiomatically biasing context. In contrast, in literally biasing context sentences, in which an idiomatic interpretation is absent, we hypothesized the semantics of the target word to be strongly activated leading to co-activation of semantically related words, resulting in a faster naming response to the semantically related target word (REL) as opposed to the unrelated target word (UNREL). These findings would

suggest that during the processing of opaque idiomatic expressions, the idiom-final word is not activated at the semantic level. However, the semantics of this target word is activated if it is embedded in a literally biasing context, even when using a word-naming task that focuses on orthography.

However, because the idiom-final word's form characteristics must be retrieved to integrate it successfully into the idiomatic context, we hypothesized that we would observe activation of the idiom-final noun at the orthographic level in terms of word frequency. More specifically, higher individual word frequencies were predicted to be associated with slower naming latencies in the idiomatic context, which would be in line with van Ginkel and Dijkstra (2019).

## 5.2 Experiment 1

### 5.2.1 Methods

#### *Participants*

Thirty-two native speakers of Dutch participated in the first experiment (24 females and 8 males). They were between 19 and 33 years old ( $M = 23.7$ ;  $SD = 3.63$ ), and had a normal or corrected-to-normal vision. They received compensation for participation in terms of a gift card or participant credits. This study was ethically assessed and approved by the Ethics Assessment Committee (EAC) of the Faculty of Arts of Radboud University Nijmegen (number 3382).

#### *Materials and design*

The materials consisted of 180 experimental sentences (30 sets of six sentences) and 60 filler sentences. The target word was always the last word of the sentence. In the filler sentences, the target word was a noun in a literal context. In the experimental sentences, however, the target word was either a noun that was part of an idiom (idiomatic context), or the same noun embedded in a literal context. The experiment involved a within-subject design with the variables Context (Idiomatic and Literal) and Condition (COR, REL, and UNREL).

In each context (Idiomatic and Literal), three versions of the same sentence were created by changing the target word. The target word was either the expected/correct word given the context (COR), a word that



was semantically related to the expected word (REL), or a word that was semantically unrelated to the expected word (UNREL). See Table 5.1 for example stimuli.

The materials were divided in three master lists containing 210 sentences: 180 experimental sentences (90 idiomatic sentences and 90 literal sentences with COR, REL and UNREL evenly distributed; 30 sentences of each condition), and 30 filler sentences with an expected target word only. Each participant received a pseudo-randomization of one of the three lists.

Table 5.1 Example sentences of experimental items

Condition	Example sentence
<i>Idiomatic</i>	
COR	De getrainde dief liep uiteindelijk toch tegen de <u>lamp</u> <i>The trained thief eventually walked against the <u>lamp</u></i>
REL	De getrainde dief liep uiteindelijk toch tegen de <u>warmte</u> <i>The trained thief eventually walked against the <u>warmth</u></i>
UNREL	De getrainde dief liep uiteindelijk toch tegen de <u>helm</u> <i>The trained thief eventually walked against the <u>helmet</u></i>
<i>Literal</i>	
COR	Het kind kan niet slapen zonder licht van een kleine <u>lamp</u> <i>The child cannot sleep without light of a little <u>lamp</u></i>
REL	Het kind kan niet slapen zonder licht van een kleine <u>warmte</u> <i>The child cannot sleep without light of a little <u>warmth</u></i>
UNREL	Het kind kan niet slapen zonder licht van een kleine <u>helm</u> <i>The child cannot sleep without light of a little <u>helmet</u></i>

*Idiom selection.* We compiled a database of 374 Dutch idiomatic expressions that were rated by 390 native speakers of Dutch on different dimensions, such as Transparency, Familiarity, and Imageability. The ratings were found to be highly reliable (Hubers et al., 2019, 2018). We selected 30 opaque idiomatic expressions from this database as a basis for the experimental sentences. The idiomatic expressions included in

this study had a mean transparency rating of 2.22 on a scale from 1 to 5 (SD = 0.35; range = 1.31 – 2.61) and were said to be encountered quite frequently in daily life (M = 3.00; SD = 0.75; range = 2.04 – 4.76, scale 1-5).

*Target word selection.* The semantically related target words were obtained from the word association database from De Deyne and Storms (2008) if possible. If no appropriate word associations were available, we thought of semantically related words ourselves. In a pre-test, all potential REL and UNREL target words were tested with respect to their semantic relatedness to the expected target word (COR). The pre-test consisted of a rating task in which participants had to indicate to what extent word pairs were related on a 5-point Likert scale (ranging from 1 'not related at all' to 5 'highly related'). In total, 79 Dutch native speakers participated in two versions of the pre-test. We selected REL words if the average association score was above 3.5 and UNREL words if the association score was below 2.5. The REL words included in the experiment had an average association score of 4.33 (SD = 0.37; range = 3.60 – 4.93). The average association score for the UNREL words included in the experiment was 1.49 (SD = 0.35; range = 1.04 - 2.14).

Target word frequency and target word length in letters were matched across conditions. We extracted the word frequencies per million from SUBTLEX-NL (Keuleers, Brysbaert, & New, 2010). The conditions (COR, REL and UNREL) did not significantly differ in terms of log-transformed word frequency,  $F(2,87) = .055, p = 0.947$  (M = 2.75; SD = 0.64). The conditions did not significantly differ in terms of target word length,  $F(2,87) = .920, p = 0.083$  (M = 4.86; SD = 1.30).

We controlled for the initial sound of the target words, given that in word naming especially fricatives and plosives may trigger the voice key later than for example nasals, even if the articulatory onset of these phonemes takes place at the same time (e.g., Duyck et al., 2008; Tyler et al., 2005). In line with Duyck et al. (2008), we divided the target words in five categories depending on their initial phoneme: vowels, fricatives, nasals, plosives, and approximants. The target words were selected in such a way that within each condition (COR, REL, and UNREL), the phonetic categories of the initial sounds were similarly distributed, especially with respect to fricatives and plosives.

*Cloze probability.* We controlled for the cloze probability of the expected target words (COR) in both the idiomatic and literal contexts. To this end, we conducted a pre-test including 219 potential experimental sentences without the final word (the target word). These sentences were divided over two lists. Participants were asked to fill in the first word that came to mind upon reading the sentences. In total, a group of 17 participants carried out this first version of the cloze test (age  $M=20.6$ ;  $SD=1.6$ ; females = 14). A subset of the sentences was adapted and tested again. The second version of the cloze test contained both the adjusted sentences and the sentences that had been already tested. The design and procedure of this test was the same as before. In total, 38 people participated (31 females). They were on average 32.6 years old ( $SD = 12.7$ ). In a third version of the cloze test, the remaining set of 43 adapted sentences were tested by a group of 20 participants (age  $M = 31.3$ ;  $SD = 12.7$ ). The experimental sentences in both the literal and the idiomatic contexts had comparable cloze probabilities (LIT:  $M = 0.82$ ,  $SD = 0.15$ ; IDIOM:  $M = 0.83$ ;  $SD = 0.16$ ),  $t(178)=.0387$ ,  $p = .699$ .

*Sentence plausibility.* To obtain information about the plausibility of the sentences containing a violation (REL and UNREL), we carried out a sentence plausibility test. An independent group of 32 native speakers of Dutch were asked to assess whether the sentences were plausible on a scale ranging from 1 (not plausible at all) to 7 (highly plausible). All materials were divided over three lists containing 180 sentences (90 literal sentences and 90 idiomatic sentences with COR, REL, and UNREL evenly distributed). The participants were randomly assigned to the list resulting in almost evenly distributed groups of participants per list (cf. 9, 11 and 12 participants). Half of the participants in each group received the list in reversed order. Table 5.2 provides the mean plausibility ratings for the experimental sentences. The literal contexts were rated as more plausible than idiomatic contexts ( $F(1,31) = 126.82$ ;  $p < 0.01$ ). In addition, Condition ( $F(1.54, 47.60) = 1048.04$ ;  $p < 0.01$ ) and the interaction effect of Context and Condition ( $F(1.67, 51.82) = 48.63$ ;  $p < 0.01$ ) were significant. Simple effect analyses showed that COR, REL, and UNREL significantly differed from each other in both the Literal and Idiomatic context.

See Table S5 in the Supplementary Materials for the idiomatic expressions included in the experiment and their corresponding target words.

Table 5.2  
Mean Plausibility Ratings and SDs for the experimental sentences (scale 1-7)

<b>Condition</b>	<b>Context</b>	
	<b>Literal</b>	<b>Idiomatic</b>
COR	6.5 (0.5)	5.9 (0.9)
REL	3.6 (0.6)	1.9 (0.6)
UNREL	1.6 (0.3)	1.4 (0.4)

#### *Procedure*

The participants were tested in a soundproof booth. The experiment was programmed in PsychoPy (Peirce, 2007). Word naming was recorded with a head-mounted microphone (SHURE WH-20-XLR), and naming latencies were calculated by the PsychoPy voice-key module (Peirce, 2007) and checked afterwards in Matlab (MathWorks, 2016).

The experiment consisted of two parts: (1) the familiarization phase and (2) the main experiment. For the first part, participants were told to read idiom meaning pairs. Although we selected idioms for our experiment that were relatively frequent, we included a familiarization phase prior to the main experiment because we intended to conduct this experiment also with L2 learners of Dutch, which are generally less familiar with the idioms (see Chapter 6). For this group we wanted to increase the likelihood that participants recognized the idioms as such. As for the main experiment, participants were instructed that they would read sentences presented word by word on the screen with the last word of each sentence presented in red. They were asked to read aloud the red word as quickly as possible. Furthermore, participants were instructed that every now and then they would be presented with comprehension questions about the sentence directly preceding the question. They were asked to answer the question with yes or no by pressing the corresponding buttons on the button box. In this way we forced the participants to actually read the sentence context preceding the target word.

In the familiarization phase, all 30 idiomatic expressions included in the main experiment were presented to the participants along with their meanings. The idiomatic expressions were presented at the center of the screen in white on a black background with the meaning of the idioms directly below them. After 30 seconds, the next idiom-meaning pair automatically appeared on the screen. No explicit task was formulated. This part of the experiment took approximately 5 minutes.

The main experiment started with a practice phase consisting of 11 practice trials and 3 comprehension questions in order for the participants to get used to the task. After the practice phase they had the opportunity to ask questions if anything was unclear.

A trial started with a fixation cross that was presented for 500 ms, followed by a blank screen of 300 ms. Subsequently, a sentence was presented in a word-by-word fashion. The words were presented at the center of the screen in white on a black background. Each word was displayed for 300 ms, after which a blank screen was presented for 300 ms. The last word of the sentence, the target word, was presented in red and disappeared after 2500 milliseconds or when the voice-key triggered. The next trial was presented automatically 2500 milliseconds after the onset of the target word.

After the main experiment, participants filled in a background questionnaire and were tested on their knowledge of idiomatic expressions included in the experiment by means of an open-ended question about the idiom meanings. In total, it took participants 1 hour to complete the experiment.

#### *Data analysis*

We performed linear mixed effects regression analyses to analyze the naming latencies. These analyses were conducted in the statistical software package 'R' version 3.4.0 (R Development Core Team, 2008), and the R packages 'lme4' (Bates et al., 2015), 'lmerTest' (Kuznetsova et al., 2017), and 'effects' (Fox, 2003) were used. The models were built in a forward manner, starting off with a basic model including a random intercept for participants and the variables of interest (Context and Condition). Subsequently, we added different predictors to the model (random and fixed factors) one by one based on theory. After adding a predictor, we examined whether the model fit improved. If this was not

the case, we decided not to include this predictor in the model. The final model is reported in this paper.

### 5.2.2 Results

Naming errors and trials with naming latencies shorter than 360 ms were removed from the data (2.8 %). Three participants were removed because of poor performance on the comprehension questions (< 70% correct). Responses at 2.5 SDs from the mean were removed on the participant and item level (2%). The average naming latencies and SDs per Context and Condition are presented in Table 5.3.

Table 5.3  
Mean Naming Latencies and SDs in Experiment 1.

Condition	Context	
	Literal	Idiomatic
COR	579 (117)	565 (124)
REL	607 (119)	592 (116)
UNREL	614 (121)	592 (112)

We performed a linear mixed effects regression analysis to analyze the data. The log-transformed reaction times were used as the dependent variable. In our final regression model we included the following predictors as fixed effects: (1) Context (Idiomatic and Literal), (2) Condition (COR, REL, and UNREL), (3) Trial number, (4) Initial sound (Vowels, Plosives, Fricatives, Approximants, and Nasals), (5) Target word frequency (logged), (6) Idiom transparency, (7) Idiom imageability, (8) Context x Condition, (9) Context x Idiom transparency, and (10) Context x Idiom imageability.

In addition, we included target word (intercept only) and participants (intercept and random slope of Trial number) as random effects. We included target word as an item related random effect instead of idiom, because the target words occurred in both the literal and the idiomatic contexts, while the idioms were only presented as such in the idiomatic context. Reference categories for the categorical predictors are Literal (for Context), Fricatives (for Initial sound), and REL (for Condition). The variables Trial number, Idiom transparency, and Idiom imageability were mean centered. The model is presented in Table 5.4.

Table 5.4 Regression Model Experiment 1 with Logged Naming Latencies as the Dependent Variable

<b>Fixed effects</b>	<b>Beta</b>	<b>Std. Error</b>	<b>t value</b>	
(Intercept)	6.3531	0.0251	253.530	***
Trial number	0.0044	0.0043	1.012	
Initial sound (Vowels)	0.0576	0.0142	4.052	***
Initial sound (Plosives)	0.0510	0.0086	5.95	***
Initial sound (Nasals)	0.0086	0.0133	0.645	
Initial sound (Approximants)	0.0614	0.0145	4.244	***
Target word frequency	-0.0059	0.0027	-2.218	*
Context (Idiomatonic)	-0.0249	0.0069	-3.603	***
Condition (COR)	-0.0402	0.0100	-4.022	***
Condition (UNREL)	0.0177	0.0099	1.788	.
Idiom transparency	-0.0024	0.0139	-0.175	
Idiom imageability	0.0050	0.0060	0.834	
Context (Idiomatonic) x Condition (COR)	-0.0075	0.0098	-0.764	
Context (Idiomatonic) x Condition (UNREL)	-0.0115	0.0098	-1.182	
Context (Idiomatonic) x Idiom transparency	-0.0498	0.0122	-4.089	***
Context (Idiomatonic) x Idiom imageability	-0.0189	0.0059	-3.177	**
<b>Random effects</b>	<b>Variance</b>	<b>Std. Deviation</b>	<b>Corr</b>	
Target word	0.00074	0.0272		
Participant				
Intercept	0.01571	0.1254		
Trial number	0.00042	0.0206	0.27	
Residual	0.01983	0.1408		

Note. \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

The analyses revealed no significant interaction effect between Context and Condition. The differences between COR and REL ( $\beta = -0.008$ ,  $SE = 0.01$ ,  $p > .05$ ) and REL and UNREL ( $\beta = -0.012$ ,  $SE = 0.01$ ,  $p > .05$ ) in the Idiomatic and Literal contexts were similar. Naming latencies in response to the correct target word were significantly faster than to the related target words in the literal context ( $\beta = -0.040$ ,  $SE = 0.01$ ,  $p < .001$ ). Surprisingly, the naming latencies for the semantically related target words did not significantly differ from those to the unrelated target words in the literal context ( $\beta = 0.018$ ,  $SE = 0.01$ ,  $p = .076$ ).

Similar results were found for the effect of Condition in the idiomatic context. A relevelled version of the model showed significantly faster responses to the correct target words in the idiomatic context than to the semantically related target words ( $\beta = -0.050$ ,  $SE = 0.010$ ,  $p < .001$ ), and no significant differences between the semantically related and unrelated target words ( $\beta = 0.006$ ,  $SE = 0.010$ ,  $p > .05$ ). A general facilitatory effect of Target word frequency was found ( $\beta = -0.006$ ,  $SE = 0.003$ ,  $p < .05$ ): higher target word frequencies were associated with faster naming latencies.

Idiom transparency and Idiom imageability turned out to affect naming latencies in the idiomatic context only as indicated by the significant interaction effects between Context and Idiom transparency ( $\beta = -0.050$ ,  $SE = 0.012$ ,  $p < .001$ ), and Context and Idiom imageability ( $\beta = -0.019$ ,  $SE = 0.006$ ,  $p < .01$ ). A relevelled version of the model showed a facilitatory effect of Idiom transparency ( $\beta = -0.052$ ,  $SE = 0.014$ ,  $p < .001$ ), and Idiom imageability ( $\beta = -0.014$ ,  $SE = 0.006$ ,  $p < .05$ ), in the idiomatic context irrespective of Condition: The more transparent and imageable an idiom, the faster the naming latencies in response to the target word.

### 5.2.3 Discussion

We found faster naming latencies for the correct target word in the idiomatic context as opposed to the semantically unrelated target words. This shows that idiomatic expressions are recognized as such and suggests that they have a separate representation in the mental lexicon. In the literal context, a comparable facilitation effect was found for correct target words as compared to semantically unrelated target words, indicating that the correct target word is anticipated based on the literally



biasing context. Both effects are in line with the findings of Rommers et al. (2013).

In addition, no difference was observed in terms of naming latencies between the semantically related and semantically unrelated target words in the idiomatic context. More specifically, activation did not spread from the idiom-final noun to a literally related target word, suggesting that the individual word meanings are not activated, comparable to Rommers et al. (2013). However, we are unable to interpret the lack of this effect in a reliable way, because we did not observe faster naming latencies to the semantically related target words compared to the unrelated target words in the literal context. This finding is rather surprising, because several studies have shown facilitation of the semantically related target word in highly biasing literal contexts with a similar experimental paradigm using EEG or a lexical decision task (Federmeier, 2007; Federmeier & Kutas, 1999; Federmeier et al., 2002; Rommers et al., 2013). The lack of a facilitation effect in the literal context may be due to the nature of the task employed in our study, since word-naming does not require the semantics. In addition, semantic priming is known to become stronger with increased prime durations (e.g., Holcomb, Reder, Misra, & Grainger, 2005; Lee, Rayner, & Pollatsek, 1999) and longer stimulus onset asynchronies (SOAs) (e.g., Vorberg, Mattler, Heinecke, Schmidt, & Jens, 2004). Therefore, activation may not have had enough time to spread from the correct and expected target word to the semantically related words.

However, the target words seemed to be activated at the orthographic level of representation, as observed by a facilitatory effect of Target word frequency. This effect was not modulated by Context or Condition, and fits in the general finding that higher frequency words lead to faster RTs. This, however, is in contrast with Titone and Libben (2008), and van Ginkel and Dijkstra (2019), who both reported that an increase in individual word frequency (verb and final-noun frequency respectively) led to slower RTs during idiom processing.

Although no semantic effects were observed on the level of the individual words, we did find semantic effects of the idiom as whole, indicating that the word-naming task is sensitive to semantics. We observed effects of the idiom properties transparency and imageability in

the idiomatic context only. These effects were not present in the literal context, because in this context the target words were not part of the idiom. Even though we only included opaque idioms, the variation among idioms in terms of their transparency scores was large enough to lead to an effect of transparency. This facilitatory effect is in line with many studies on idiom processing (e.g., Gibbs, Nayak, & Cutting, 1989; Libben & Titone, 2008; van Ginkel & Dijkstra, 2019). If the individual words contribute to the figurative meaning, it is easier to process the idiom final noun, as opposed to if the individual words do not contribute to the figurative meaning. We also found a facilitatory effect of idiom imageability. This facilitatory effect is also in line with research on idiom learning (Boers et al., 2009; Steinel et al., 2007). However, some researchers have observed inhibitory effects of imageability on idiom processing (e.g., Cacciari & Glucksberg, 1995). In two online experiments, Cacciari and Glucksberg (1995) found that increased imageability was accompanied by increased RTs. In addition, an offline image production task revealed that people tend to visualize the meaning of the individual words composing the idiom rather than the figurative meaning of the idiom. It seems that if idioms are highly imageable and people depict the literal image, this hampers idiom processing. Imageability might therefore be related to literal plausibility, i.e. the extent to which an idiom can be interpreted literally. If an idiom can be interpreted literally, people may be more likely to visualize the literal meaning. However, for opaque idiomatic expressions, in which the individual word meanings do not really contribute to the figurative meaning, it may be rather difficult to visualize the literal meaning of the idiom. Thus, if participants associate an image to the idiomatic expressions, they might visualize the figurative meaning instead, leading to the facilitatory effect we observed in our study.

In order to properly interpret the pattern of results obtained for the idiomatic context, we must establish whether there can be a facilitation effect of the semantically related target word in the literal context. Therefore, we conducted a second experiment in which we delayed the presentation of the target word. This delay would give the target word's activation more time to spread and thus increases the

chance of observing a facilitation effect of the semantically related target word in the literal context.

## 5.3 Experiment 2

### 5.3.1 Methods

#### *Participants*

In total, 29 native speakers of Dutch participated in the experiment (22 females, and 7 males). They were between 18 and 46 years old ( $M = 24.03$ ,  $SD = 6.78$ ), and had a normal or corrected-to-normal vision. They received compensation for participation in terms of a giftcard or participant credits.

#### *Materials and design*

The same materials and design were used as in Experiment 1.

#### *Procedure*

Almost the same procedure was used as in Experiment 1. The experiment consisted of two parts: a familiarization phase and the main experiment. Experiment 2 differed from Experiment 1 with respect to the presentation of the target words in the main experiment. Similar to Experiment 1, sentences were presented visually in a word-by-word fashion presenting each word for 300 ms followed by a blank screen for 300 ms. However, the target word was not presented after a 300 ms blank screen, as in Experiment 1, but instead was delayed and displayed after a 500 ms blank screen.

#### *Data analysis*

The same procedure was used to analyze the data as in Experiment 1.

### 5.3.2 Results

Naming errors and trials with naming latencies shorter than 360 ms and longer than 1333 ms were removed from the data (7.0 %). Three participants were removed because of poor performance on the comprehension questions (< 70% correct). Responses at 2.5 SDs from the mean were removed on the participant and item level (2.1%). The average naming latencies and SDs per Context and Condition are presented in Table 5.5.

Table 5.5  
Mean Naming Latencies and SDs in Experiment 2.

Condition	Context	
	Literal	Idiomatic
COR	542 (113)	531 (105)
REL	568 (105)	565 (98)
UNREL	585 (113)	566 (101)

We analyzed the naming latencies by means of a linear mixed effects regression analysis with the logged naming latencies as the dependent variable. The final model consisted of the following fixed factors: (1) Trial number (mean centered), (2) Initial sound (reference category: Fricatives), (3) Target word length, (4) Target word frequency (logged and mean centered), (5) Context (reference category: Literal), (6) Condition (reference category: REL), (7) Idiom imageability (mean centered), (8) Context x Condition, (9) Context x Target word frequency, (10) Condition x Target word frequency, (11) Context x idiom imageability, (12) Context x Condition x Target word frequency. As random effects we included Participant (intercept and random slope of Trial number) and Target word (intercept only). The model is presented in Table 5.6.

This analysis revealed an interesting significant three-way interaction with Target word frequency, Context and Condition. More specifically, the effect of Target word frequency on naming latencies was different for the correct target word as opposed to the semantically related target word in the idiomatic context, but not in the literal context ( $\beta = 0.029$ ,  $SE = 0.007$ ,  $p < .001$ ). The interaction effect is visualized in Figure 5.1.

A significant interaction effect between Context and idiom imageability was observed, indicating that idiom imageability significantly affected naming latencies in the idiomatic context, but not in the literal context ( $\beta = -0.019$ ,  $SE = 0.006$ ,  $p < .01$ ). A relevelled version of the model showed a facilitatory effect of idiom imageability on naming latencies in the idiomatic context ( $\beta = -0.017$ ,  $SE = 0.005$ ,  $p < .01$ ). Adding other idiom properties in interaction with Context did not significantly affect the naming latencies, as this did not lead to an improved model fit.

Table 5.6 Regression Model Experiment 2 with Logged Naming Latencies as the Dependent Variable

<b>Fixed effects</b>	<b>Beta</b>	<b>Std. Error</b>	<b>t value</b>
(Intercept)	6.2620	0.0275	227.968 ***
Trial number	-0.0045	0.0043	-1.062
Initial sound (Vowels)	0.0533	0.0124	4.301 ***
Initial sound (Plosives)	0.0389	0.0071	5.469 ***
Initial sound (Nasals)	0.0078	0.0115	0.683
Initial sound (Approximants)	0.0073	0.0116	0.628
Target word length	0.0087	0.0024	3.602 ***
Target word frequency (TW freq)	-0.0029	0.0039	-0.744
Context (Idiomatic)	-0.0060	0.0068	-0.885
Condition (COR)	-0.0479	0.0089	-5.41 ***
Condition (UNREL)	0.0296	0.0088	3.349 ***
Idiom imageability	0.0023	0.0052	0.437
Context (Idiomatic) x Condition (COR)	-0.0170	0.0097	-1.746 .
Context (Idiomatic) x Condition (UNREL)	-0.0224	0.0097	-2.311 *
Context (Idiomatic) x Target word frequency	-0.0095	0.0042	-2.247 *
Condition (COR) x Target word frequency	-0.0087	0.0063	-1.39
Condition (UNREL) x Target word frequency	0.0000	0.0058	0.006
Context (Idiomatic) x Idiom imageability	-0.0188	0.0058	-3.234 **
Context (Idiomatic) x Condition (COR) x TW freq	0.0286	0.0068	4.189 ***
Context (Idiomatic) x Condition (UNREL) x TW freq	0.0037	0.0063	0.596
<b>Random effects</b>	<b>Variance</b>	<b>Std. Deviation</b>	<b>Corr</b>
Target word	0.0005	0.0211	
Participant	0.0141	0.1188	
Trial number	0.0004	0.0191	-0.09
Residual	0.0166	0.1288	

Note. \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

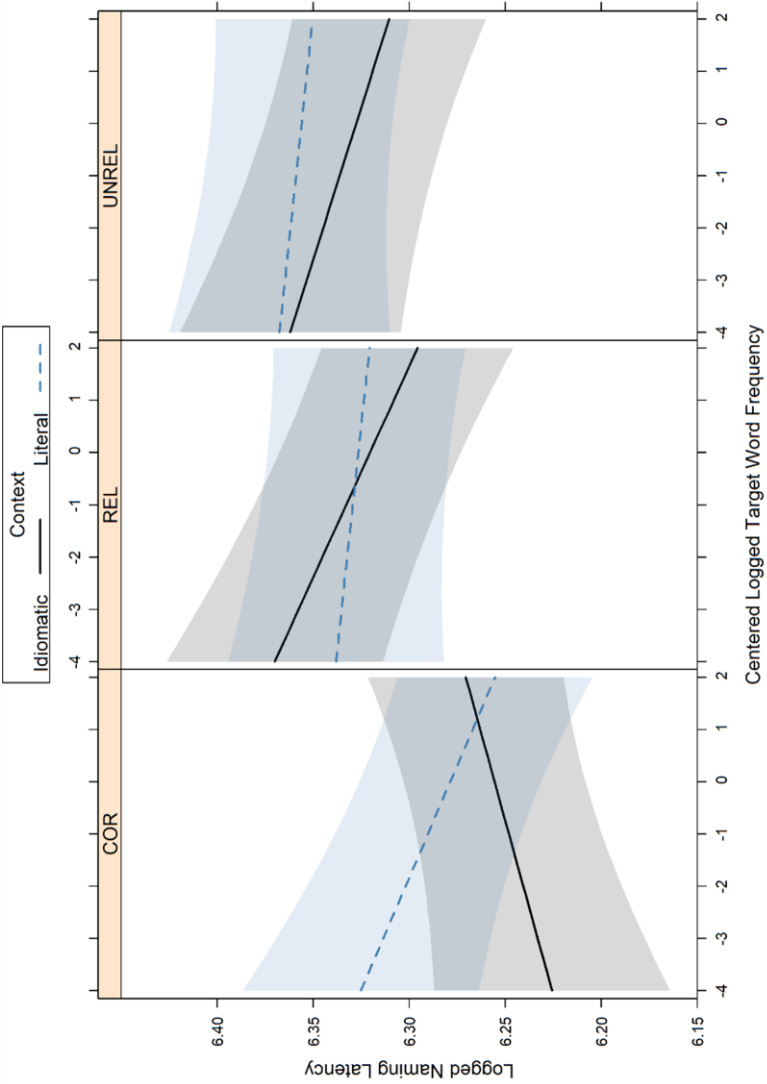


Figure 5.1 The effect of target word frequency by context and condition. The error bands are based on SEs.

*Separate analyses*

In order to obtain a better insight in the three-way interaction effect, we analyzed the idiomatic and literal contexts separately. For both sub-analyses, a linear mixed effects regression analysis was carried out including the same random and fixed factors as in the regression model based on the complete dataset except Context.

The following fixed factors were included: (1) Trial Number, (2) Initial Sound (reference category: Fricatives), (3) Target word length, (4) Target word frequency (logged and mean centered), (5) Condition (reference category: REL), (6) Idiom imageability (mean centered) (7) Target Word Frequency x Condition. As random effects we included Participant (intercept and random slope of Trial number) and Target word (intercept only). The regression models based on the Literal and Idiomatic Context Sentences are presented in Table 5.7, and Table 5.8 respectively.

In the analysis based on the Literal context sentences only, we found no significant interaction effect between Condition and Target word frequency ( $\beta = -0.009$ ,  $SE = 0.006$ ,  $p > .05$ , and  $\beta = -0.001$ ,  $SE = 0.006$ ,  $p > .05$ ). However, a facilitatory effect of Target word frequency was observed for correct target words (relevelled version of the model:  $\beta = -0.011$ ,  $SE = 0.005$ ,  $p < .05$ ). Crucially, the analysis revealed a significant differences between COR, REL, and UNREL. Participants were significantly slower in response to semantically related target words than to their correct counterparts ( $\beta = -0.047$ ,  $SE = 0.009$ ,  $p < .001$ ), but faster than in response to the semantically unrelated target words ( $\beta = 0.030$ ,  $SE = 0.009$ ,  $p < .001$ ). Idiom imageability did not significantly affect naming latencies in the Literal context sentences ( $\beta = 0.002$ ,  $SE = 0.005$ ,  $p > .05$ ). In addition, significant covariates were Target word length and Initial Sound. The longer the target words, the longer the naming latencies ( $\beta = 0.009$ ,  $SE = 0.003$ ,  $p < .01$ ), and target words starting with a fricative were named faster than target words starting with a vowel ( $\beta = 0.061$ ,  $SE = 0.014$ ,  $p < .001$ ) and a plosive ( $\beta = 0.048$ ,  $SE = 0.008$ ,  $p < .001$ ).

Table 5.7 Regression Model Experiment 2 for the Literal Context Sentences only with Logged Naming Latencies as the Dependent Variable

<b>Fixed effects</b>	<b>Beta</b>	<b>Std. Error</b>	<b>t value</b>
(Intercept)	6.2591	0.0298	210.093 ***
Trial number	-0.0061	0.0048	-1.271
Initial sound (Vowels)	0.0606	0.0143	4.243 ***
Initial sound (Plosives)	0.0475	0.0082	5.801 ***
Initial sound (Nasals)	-0.0014	0.0132	-0.11
Initial sound (Approximants)	0.0141	0.0134	1.054
Target word length	0.0086	0.0028	3.073 **
Target word frequency	-0.0024	0.0038	-0.642
Condition (COR)	-0.0468	0.0086	-5.451 ***
Condition (UNREL)	0.0298	0.0085	3.488 ***
Idiom imageability	0.0015	0.0051	0.290
Condition (COR) x Target word frequency	-0.0089	0.0061	-1.458
Condition (UNREL) x Target word frequency	-0.0010	0.0056	-0.186
<b>Random effects</b>	<b>Variance</b>	<b>Std. Deviation</b>	<b>Corr</b>
Target word	Intercept	0.0003	0.0187
Participant	Intercept	0.0162	0.1272
	Trial number	0.0004	0.0198
Residual		0.0169	0.09

Note. \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .



The regression model based on the Idiomatic context sentences only (see Table 5.8) revealed a significant interaction effect between Condition and Target word frequency. The effect of Target word frequency was significantly different for correct words as opposed to semantically related words ( $\beta = 0.020$ ,  $SE = 0.007$ ,  $p < .01$ ) and semantically unrelated target words (releveled version of the model:  $\beta = 0.015$ ,  $SE = 0.007$ ,  $p < .05$ ). The effect of Target word frequency on naming latencies was similar for semantically related and unrelated target words ( $\beta = 0.004$ ,  $SE = 0.006$ ,  $p > .05$ ). Interestingly, naming latencies for correct target words were significantly faster than for semantically related target words ( $\beta = -0.066$ ,  $SE = 0.010$ ,  $p < .001$ ), whereas naming latencies for semantically related and unrelated target words did not differ ( $\beta = 0.006$ ,  $SE = 0.010$ ,  $p > .05$ ). A releveled version of the model showed that naming latencies for correct target words were also significantly faster than naming latencies for semantically unrelated target words ( $\beta = -0.071$ ,  $SE = 0.010$ ,  $p < .001$ ). Idiom imageability significantly influenced naming latencies: the more imageable an idiom, the faster the naming latencies ( $\beta = -0.015$ ,  $SE = 0.006$ ,  $p < .01$ ). Furthermore, Target word length influenced naming latencies: The longer the target word, the longer the naming latency ( $\beta = 0.009$ ,  $SE = 0.003$ ,  $p < .01$ ). In addition, naming latencies in response to target words starting with a fricative were different from naming latencies in response to target words starting with a vowel ( $\beta = 0.046$ ,  $SE = 0.016$ ,  $p < .01$ ) or a plosive ( $\beta = 0.030$ ,  $SE = 0.010$ ,  $p < .01$ ).

Table 5.8 Regression Model Experiment 2 for the Idiomatic Context Sentences only with Logged Naming Latencies as the Dependent Variable

	<b>Beta</b>	<b>Std. Error</b>	<b>t value</b>
<b>Fixed effects</b>			
(Intercept)	6.2606	0.0284	220.274 ***
Trial number	-0.0029	0.0047	-0.625
Initial sound (Vowels)	0.0463	0.0162	2.863 **
Initial sound (Plosives)	0.0302	0.0093	3.251 **
Initial sound (Nasals)	0.0148	0.0149	0.991
Initial sound (Approximants)	0.0007	0.0150	0.045
Target word length	0.0089	0.0032	2.800 **
Target word frequency	-0.0126	0.0044	-2.893 **
Condition (COR)	-0.0663	0.0097	-6.848 ***
Condition (UNREL)	0.0055	0.0096	0.575
Idiom imageability	-0.0151	0.0057	-2.644 **
Condition (COR) x Target word frequency	0.0197	0.0069	2.856 **
Condition (UNREL) x Target word frequency	0.0044	0.0063	0.694
<b>Random effects</b>	<b>Variance</b>	<b>Std. Deviation</b>	<b>Corr</b>
Target word	0.0007	0.0258	
Participant	0.0120	0.1098	
	0.0004	0.0190	-0.27
Residual	0.0160	0.1266	

Note. \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

### 5.3.3 Discussion

In line with the first experiment, we found that participants responded faster to the correct target word as compared to unrelated target words in both the literal and the idiomatic context. The facilitation effect suggests that participants used the context to anticipate the final word of the sentence. In the idiomatic context, this shows that participants recognized the idioms as such.

Importantly, we observed faster naming latencies for the semantically related word as compared to the unrelated word in the literal context, which is in line with earlier studies using this paradigm with lexical decision and EEG (Federmeier, 2007; Federmeier & Kutas, 1999; Federmeier et al., 2002; Rommers et al., 2013). A delayed target word presentation of 200 ms was enough to increase the activation of the correct target word to such an extent that it was able to spread to words that are semantically related to this word.

In the idiomatic context, we found no facilitation of the semantically related target words compared to the unrelated words. This, in combination with the presence of this effect in the literal context, suggests that the literal word meanings were not activated during the processing of opaque Dutch idioms. In other words, the facilitatory effect due to semantic relatedness in the idiomatic context was probably reduced because the individual word meanings were suppressed. This is in line with the findings of Rommers et al. (2013) who used the same paradigm with lexical decision and EEG. However, we did find activation of the idiom-final noun at the orthographic level in terms of individual word frequency. More specifically, higher target word frequencies were associated with slower naming latencies of the idiom-final noun in the idiomatic context, while in the literal context higher target word frequencies did not lead to faster naming latencies for the correct target word. During idiom processing higher individual word frequency seems to hinder idiom processing. This inhibitory effect on idiom processing is in line with the verb frequency effect reported by Libben and Titone (2008), and with van Ginkel and Dijkstra (2019), who observed a comparable effect of the idiom-final noun frequency. This suggests that, although the idiom-final noun is not activated strongly enough for the activation to spread to semantically related words, even in a strongly

idiomatically biasing context containing opaque idioms, participants still access aspects of this word related to its literal use.

Considering the semantic effects on the level of the idiom as a whole, only idiom imageability affected naming latencies. As in Experiment 1, idiom imageability negatively affected naming latencies: the more imageable an idiom, the faster the naming latencies. No effects were found for idiom transparency or familiarity.

#### **5.4 General Discussion**

In the present studies, we examined the role of the individual words during the processing of opaque idiomatic expressions at the semantic and the orthographic levels. To this end, we adopted the same paradigm as in Rommers et al. (2013), but used word-naming instead of lexical decision, a task focusing on orthography. Based on Rommers et al. (2013), we hypothesized that the expected target word in the idiomatic context (the idiom-final noun) would not be activated at the semantic level of representation. However, because the word form of the idiom-final noun must be identified to verify that the idiom is actually presented, we did expect to find effects of the target word in the idiomatic context at the orthographic level in terms of its frequency.

In the first experiment, a facilitatory effect of target word frequency was found that was not modulated by Context or Condition, but indicated that the target words were activated at the level of the orthography. Moreover, participants responded faster to the correct target word in both the idiomatic and the literal contexts as compared the unrelated target word, suggesting that the target words were anticipated and that the idioms were recognized as such by the participants. In line with Rommers et al. (2013), no differences were found in terms of naming latencies for semantically related and unrelated target words in the idiomatic context. Surprisingly, however, we did not find the graded pattern of results that was reported by Rommers et al. (2013) in the literal context. Instead, naming latencies in response to the semantically related target word were similar to those in response to unrelated target words. While no semantic effects were observed on the level of the target words, we did find effects related to the semantics of the idiom as a whole.

Idiom transparency and imageability turned out to affect idiom processing. In line with previous studies, more transparent idioms led to faster RTs as opposed to less transparent idioms (Gibbs, Nayak, & Cutting, 1989; Libben & Titone, 2008; van Ginkel & Dijkstra, 2019). As for imageability, we observed a facilitatory effect: Imageable idioms were associated with faster naming latencies than non-imageable idioms. This finding is different from that of Cacciari and Glucksberg (1995), who showed exactly the opposite. They found that participants more often visualize the literal meaning of the idiom than the figurative interpretation. In our experiment, participants may have visualized the figurative meaning more often, because we included opaque idiomatic expressions only, leading to a facilitatory effect on idiom processing.

In the second experiment, in which we delayed the presentation of the target word by 200 ms, we did observe a graded pattern for target word condition (COR < REL < UNREL) in the literal context, indicating pre-activation of the correct word, leading to spreading activation to semantically related words. In the idiomatic context, we found that the idiom-final noun (the correct target word) was anticipated, as reflected by faster naming latencies as opposed to unrelated target words. Moreover, no difference was found between naming latencies for the semantically related and unrelated words, suggesting that the idiom-final nouns are not activated at the semantic level. Facilitatory effects were reduced or cancelled out by suppression of the individual word meanings. Although the idiom-final noun seemed to be suppressed at the semantic level, activation was observed at the orthographic level of representation, as indicated by the effect of target word frequency. This is in line with our hypothesis that for the idiomatic expression to be selected from the mental lexicon, the idiom-final noun needs to be verified. As a consequence, it should be activated at least at the form level. More specifically, the direction of the effect indicated a competition of the idiom-final noun and the idiom as a whole. The time it takes to verify whether the idiom-final noun is part of the idiomatic expression depends on its word frequency: Higher idiom-final noun frequencies lead to slower naming latencies. Although this inhibitory effect of word frequency is in contrast with the robust findings in literal language processing that higher word frequencies are associated with faster

processing times (see Brysbaert, Mandera, & Keuleers, 2018, for a review), it is in line with the limited number of studies investigating the role of single word frequency in idiom processing (Libben & Titone, 2008; van Ginkel & Dijkstra, 2019). As for the effects related to the semantics of the idiom as a whole, idiom imageability turned out to affect idiom processing in the same way as in Experiment 1. However, in contrast to Experiment 1, idiom transparency did not influence idiom processing after the presentation of the target word had been delayed.

To summarize, the results of this study show that the idiom-final noun of opaque idiomatic expressions in highly idiomatically biasing contexts is in competition with the idiom as a whole on different levels of representation. At the semantic level, activation of individual word is suppressed, since it interferes with the meaning of the idiom as a whole. At the orthographic level, however, the individual word is activated, as borne out by effects of word frequency. These results argue against both purely compositional and non-compositional models of idiom processing. According to compositional models, the individual word meanings are accessed and combined in order to retrieve the figurative meaning (Cacciari & Glucksberg, 1991; Cacciari & Tabossi, 1988; Gibbs, Nayak, Bolton, et al., 1989; Nunberg, 1979). In the current study, however, the individual word meanings are not accessed at the semantic level, since facilitation of the semantically related word is absent in the idiomatic context. Non-compositional models argue that idioms are stored as a whole in the mental lexicon and that individual word meanings are not activated during processing (Bobrow & Bell, 1973; Gibbs, 1980; Swinney & Cutler, 1979). Our study shows that this is not the case either. Although individual words are suppressed at the semantic level, we do find traces of activation at the word form level, as showed by word frequency effects.

Instead, our results provide evidence in favor of a hybrid model of idiom processing. Idiomatic expressions seem to be stored in the mental lexicon as a whole, because the idiom final word is anticipated and the individual words seem to be suppressed at the semantic level. However, even in the context of opaque idiomatic expressions embedded in a strongly idiomatically biasing context, traces of individual word activation are found at the orthographic level in terms of word frequency. This suggests that figurative and literal processing run in parallel. This is

in line with the hybrid model by Sprenger et al. (2006) when it is applied to idiom comprehension. According to this model, the idiom has a separate representation (*superlemma*) that is connected to its corresponding idiomatic meaning on the one hand, and to simple word lemmas on the other hand. The superlemmas can be accessed by activating these simple lemmas. The superlemma, in turn, activates the corresponding idiom meaning representation. In the context of the current study, the simple word lemmas have to be activated, because for the incoming target word it needs to be checked whether it is part of an idiom. However, the corresponding concepts can be ignored, because of the opacity of the idiomatic expressions included in this study, i.e. the individual word meanings do not contribute to the figurative meaning. This effect is probably strengthened by the highly idiomatically biasing context in which the idioms have been presented. Therefore, at the semantic level, the individual words are not activated, while activation is observed at the orthographic level in terms of word frequency.

This reasoning is also in line with a notion advanced by van Ginkel and Dijkstra (2019). They argue that in a word-by-word presentation of an idiomatic sentence, the figurative meaning representation will build up over time, as more information becomes available. The representation will be completed once the last word is presented. This completion process only requires the word form, not the word meaning. Our findings confirm this idea. In a strongly idiomatically biasing context containing opaque idiomatic expressions, the word meaning of the idiom-final noun is suppressed, because it does not contribute to the figurative meaning representation. However, the word form needs to be checked, which results in activation of the word form as confirmed by a word frequency effect. More specifically, the idiom-final noun is in competition with the idiom as a whole at the orthographic level. Higher idiom-final noun frequencies lead to more difficulties in integrating the idiom-final noun into the idiomatic context.

Although the semantics of the idiom-final noun was not activated, semantic effects were present at the level of the idiom. Idiom processing turned out to be affected by idiom transparency and idiom imageability in Experiment 1, while the effect of idiom transparency disappeared after the presentation of the target word was delayed (Experiment 2). This

suggests that different idiom properties seem to play a role at different points in time, which is in line with the Constraint-Based Model of idiom processing (Libben & Titone, 2008; Titone et al., 2015; Titone & Libben, 2014). Titone and Libben (2014) reported early facilitatory effects for literal plausability and familiarity. In our study, we found a similar effect for idiom imageability, which may be related to literal plausability (see Hubers et al., 2019). However, our findings with respect to transparency are somewhat different from those of Titone and Libben (2014). They reported an inhibitory effect of semantic decomposability, a dimension related to transparency, only 1000 ms post idiom offset. In contrast, we observed a facilitatory effect of transparency only in Experiment 1. After delaying the presentation of the target word, this effect disappeared.

Crucial differences between our study and that of Titone and Libben (2014) lie in the task and paradigm used. Titone and Libben (2014) used a cross-modal priming experiment in which they measured RTs on a visually presented target word that was not part of the prime sentence. In our experiment, however, the target word had to be integrated into the sentence context, since it was the final word of the sentence. As van Ginkel and Dijkstra (2019) suggest, it is difficult to compare the results of empirical studies due to task differences. Therefore, more research is needed that systematically investigates task effects on idiom processing.

## 5.5 Conclusions

The present study investigated the role of individual words during idiom processing. More specifically, we examined to what extent the individual words are activated at the semantic and the orthographic levels during the processing of opaque Dutch idiomatic expressions in highly idiomatically biasing contexts using a word-naming task focusing on orthography more than on semantics.

Our results show that the individual words are in competition with the idiom as whole at both the semantic and the orthographic levels of representation. At the meaning level, individual words were not activated, because in highly idiomatically biasing contexts containing opaque idiomatic expressions, the semantics of the idiom-final noun only



interferes with the idiom as a whole. However, even though strong contexts were used, activation of the idiom-final noun was observed at the orthographic level in terms of word frequency, reflecting a competition process between the individual word and the idiom as a whole. Higher idiom-final word frequencies were associated with slower naming latencies. In addition, although no semantic effects were observed of the individual words in the idiomatic context using word-naming, we did find semantic effects related to the idiom as a whole, showing that word-naming is sensitive to semantic effects in idiomatic contexts. Imageability and Transparency both facilitated idiom processing. Time-course seems to be important, since the effect of Transparency disappeared when the target word was presented later in time.

These results provide evidence in favor of a hybrid model of idiom processing in which the individual words and the idiom as a whole are in interaction at both the orthographic and semantic levels of representation. In an idiomatic context in which the semantics of the individual words do not contribute to the idiomatic meaning, the activation of the individual word meanings may be suppressed. However, the idiom-final noun is activated at the orthographic level, because to verify that an idiom is actually being presented, the idiom-final word form has to be identified.



## Chapter 6

### **Idiom processing by L2 learners: The role of individual words and repeated exposure**

### **Abstract**

L2 learners tend to rely more on the individual word meanings during idiom processing than native speakers (Abel, 2003; Cieśllicka, 2006; Kecskes, 2007). Some researchers argue that L2 learners approach idiomatic expressions in a fundamentally different way than native speakers (Cieśllicka, 2006), while others think L2 learners become more native-like after repeated exposure (Kecskes, 2007).

The current study investigated to what extent individual words are activated during L2 idiom processing and to what extent changes in L2 idiom processing could be observed as a function of repeated practice or exposure. To study this in depth, we combined research techniques from applied linguistics and psycholinguistics. We conducted a primed word-naming experiment with L2 German learners of Dutch after they had participated in a Computer Assisted Language Learning (CALL) experiment aimed at increasing their exposure to and experience with L2 idioms. As part of this CALL experiment, we manipulated the intensity of their practice. This allowed us to determine to what extent effects of exposure arose in the word-naming experiment. We compared the results obtained from L2 learners with the results from native speakers reported in the previous Chapter.

Our results suggest that, like native speakers, L2 learners can access the figurative meaning when they process opaque idiomatic expressions embedded in highly biasing contexts, and that the ease with which this is done is influenced by repeated exposure. The findings indicate that the idiom's component words are activated at the orthographic level, but they leave open to what extent they are also semantically activated.

### **Acknowledgments**

We are grateful to our fellow members of the ISLA team, Wendy van Ginkel, Louis ten Bosch, and Alexander Wahl for their valuable input. This work is part of the research program Free Competition in the Humanities with project number 360-70-510 NWO ISLA, which is financed by the Dutch Research Council (NWO).

## 6.1 Introduction

In the previous chapter, we investigated the effect of individual word frequency on native speaker idiom processing. We saw that native speakers access the individual words composing an idiom even when they process opaque idiomatic expressions embedded in highly idiomatically biasing contexts. Activation of the idiom-final noun was observed at the orthographic level, which was signaled by a word frequency effect. Idiom-final nouns with higher word frequency were responded to more slowly than idiom-final nouns with lower word frequency. This effect may reflect a competition process between the individual word and the idiom as a whole. At the semantic level, however, the idiom-final noun was not activated. We argued that in highly idiomatically biasing contexts containing opaque idiomatic expressions, the semantics of the idiom-final noun only interferes with the meaning representation of the idiom as a whole, and therefore is suppressed. However, the idiom-final noun is apparently orthographically activated, because the word form of the idiom-final noun is required to verify that the item indeed completes an idiom and that it is not the final word in any literal word sequence.

The effects observed above indicate a competition process between the representations of the individual words and the idiom as a whole in native speakers. But what is the role of the individual words during the processing of idiomatic expressions by L2 learners? Researchers have suggested that during L2 idiom comprehension, L2 learners rely more on individual words than native speakers do (Cieślicka, 2006; Kecskes, 2006; Wray, 2002), but Kecskes (2006) has argued that this may become less so after repeated exposure.

In sum, if L2 learners rely more on individual words than native speakers do, this may become visible in the processing of idiomatic expressions. While in native speakers the activation of the idiom's component words may be suppressed at the semantic level, this may be different for L2 learners. In addition, through repeated exposure, idiom processing by L2 learners may become more similar to that of native speakers.

The study reported on in the current chapter investigated the effects of individual word activation and exposure on the processing of opaque idiomatic expressions by L2 learners by combining techniques from applied linguistics and psycholinguistics. We investigated idiom processing in German L2 learners of Dutch after they participate in a Computer Assisted Language Learning (CALL) experiment aimed at increasing their experience with L2 idioms. In addition, we manipulated intensity of practice (see Chapter 4) to be able to determine to what extent L2 idiom processing is affected by increased experience. In the processing experiment reported in this chapter, we applied the same primed word naming paradigm for L2 learners as in Chapter 5 for Dutch native speakers. Half of the idioms included in the word naming experiment were intensively practiced during the CALL-based training session, while the other half only received limited practice. In this way, we investigated the role of individual words during L2 idiom processing and of changes in L2 idiom processing as a function of repeated practice and increased experience with L2 idioms.

To our knowledge, no studies have so far investigated the effect of repeated exposure on L2 idiom processing by combining focused training with psycholinguistic experimentation. Before zooming in on the present study, we will discuss the limited research that considered the role of individual words during the processing of idiomatic expressions by L2 learners.

### **6.1.1 Research background**

In contrast to the presence of intensive research on the role of individual words in L1 idiom processing (see Chapter 5 for an overview and related models), only a few studies have addressed this issue for L2 learners (Beck & Weber, 2016a; Cieślicka, 2006; van Ginkel & Dijkstra, 2019).

Cieślicka (2006) investigated L2 idiom processing in a cross-modal priming experiment with advanced Polish learners of English. Participants were auditorily presented with an English sentence containing an idiom. During sentence processing, participants had to decide if an upcoming target word was an existing English word or not. Advanced L2 learners responded faster if the target word was related to the literal meaning of the idiom rather than its figurative meaning. The

processing advantage of literal target words over figurative target words was taken as evidence in favor of Cieśllicka's Literal Salience Model. This model is based on the graded salience framework by Giora (1997), which suggests that salient meanings are accessed first and are more strongly activated than non-salient meanings. The Literal Salience Model holds that, in contrast to native speakers, for L2 learners the meanings of the idiom's component words are more salient than the figurative meaning. Cieśllicka (2006) argued that increased proficiency, and repeated exposure to the L2 in general and idiomatic expressions in particular, do not change the more salient status for the literal meanings of the L2 idioms. According to this account L1 idiom processing is fundamentally different from L2 idiom processing. Beck and Weber (2016) pointed out that the study by Cieśllicka (2006) did not include native speaker participants for comparison. Instead, Cieśllicka based her assumption that idiom processing in the L1 and the L2 are fundamentally different on studies investigating L1 idiom processing only (see Beck and Weber, 2016).

Beck and Weber (2016) themselves investigated idiom processing by both L2 learners, and native speakers using a similar cross-modal priming paradigm as Cieśllicka (2006). Participants were presented with auditory sentences that included idiomatic expressions. Following the auditory sentence, a target word was displayed on the screen on which participants made an English lexical decision. The target word was either figuratively related to the meaning of the idiom (FIG), semantically related to the idiom-final word (LIT), or unrelated to both the idiom and the idiom-final noun (UNREL). Both American English native speakers and highly proficient German L2 learners of English showed priming effects for FIG and LIT target words relative to unrelated target words. The target word semantically related the idiom-final word was responded to the fastest. This finding seemed to support the idea of Cieśllicka (2006) that the individual word meanings are more salient than the meaning of the idiom as a whole. However, Beck and Weber (2016) also found this effect for native speakers, for whom the idiom meaning is often thought to be more salient than the meanings of the individual words (Giora, 2002). They concluded that L2 idiom processing is very similar to L1 idiom processing and that highly proficient L2 learners not

only have access to the figurative meaning of L2 idioms, but also to the meanings of individual component words.

A study by van Ginkel and Dijkstra (2019) also compared idiom processing in L2 learners to that of native speakers in a priming study, using a similar procedure to Beck and Weber's (2016). In a lexical decision task, both Dutch native speakers and highly proficient German L2 learners of Dutch responded faster to figuratively and literally related target words than unrelated target words. However, they also reported differences between the two groups. Native speakers, but not L2 learners, were sensitive to idiom transparency and literal plausibility. More specifically, the more transparent an idiom, the faster native speakers responded to the figuratively related target words. However, if the transparent idiom also had a literally highly plausible interpretation, native speakers became slower in response to the figuratively related target word than if the transparent idiom had a literally less plausible interpretation. This suggested that integration of the figurative meaning of the idiom was hindered if the idiom is transparent and has a plausible literal interpretation, pointing at a competition process between the figurative and literal meaning of the idiom. In addition, for native speakers, a higher frequency of the idiom-final word was associated with a slower response to the literally related target word, while a facilitatory effect of word frequency was found for L2 learners. Again, this inhibitory effect observed in native speakers suggests a competition process between the idiom as a whole and its component words (see Chapter 5 for an elaborate discussion on this topic).

Van Ginkel and Dijkstra (2019) showed that, although L1 and L2 idiom processing are similar on a global level, native speakers are sensitive to more aspects of the idiom (idiom transparency and literal plausibility) and the individual words (word frequency) than L2 learners. They argue that this higher sensitivity is a matter of exposure. Because L2 learners are less exposed to the Dutch language as compared to native speakers, they have weaker representations of both the individual component words and the idiom as a whole, which makes them less sensitive to idiom properties and aspects of the individual words.

To sum up, all studies discussed above involved a similar cross-modal priming paradigm to investigate the processing of idiomatic



expressions by L2 learners. In addition, the participants included in these studies were all highly proficient L2 learners. However, the findings obtained were rather mixed. Cieślicka (2006) reported a processing advantage for literal target words as opposed to figurative targets, indicating that the individual words are activated during idiom processing. She took these findings as evidence in favor of her Literal Salience Model, assuming that for L2 learners the idiom's component words are more salient than the figurative meaning. According to the model, the salient status of the literal word meanings does not change as a function of exposure or proficiency, suggesting a fundamental difference in how L1 and L2 speakers process idiomatic expressions. Although both Beck and Weber (2016) and van Ginkel and Dijkstra (2019) found that the individual words were activated during idiom processing, they also reported a processing advantage for figurative target words as compared to unrelated targets during L2 idiom processing. In addition, the native participants in these studies showed the same pattern of results. Both Beck and Weber (2016) and van Ginkel and Dijkstra (2019) concluded that L1 and L2 idiom processing are similar. However, van Ginkel and Dijkstra (2019) argued that the subtle differences between L1 and L2 idiom processing as observed in their study, may have been due to the L2 learners' limited exposure to the L2 and L2 idioms. However, to our knowledge, the effect of exposure to L2 idioms on L2 idiom processing has not been systematically studied so far.

### **6.1.2 Current study**

To gain more insight into the role of exposure in L2 idiom processing, we investigated to what extent the individual words are activated during the processing of L2 opaque idiomatic expressions by combining techniques from applied linguistics and psycholinguistics.

German L2 learners of Dutch participated in the word naming experiment described in Chapter 5 after taking part in the CALL-based learning study reported in Chapter 4. During the CALL-based learning study, we manipulated Intensity of Practice in such a way that half of the idioms included in the learning phase received limited practice (2 presentations), while the other half received intensive practice (12 presentations) (see Chapter 4 for more details). The idiomatic

expressions included in the word naming experiment were also part of the learning study either receiving intensive or limited practice.

The word naming experiment was inspired by the study of Rommers, Dijkstra, and Bastiaansen (2013) and was also used in the previous chapter (Chapter 5) to study the activation of the individual words during idiom processing by native speakers. This paradigm is different from the cross-modal priming technique adopted by Beck and Weber (2016), Cieślicka (2006), and van Ginkel and Dijkstra (2019), because the activation of the idiom and the individual words are measured at a different point in time. By adopting the paradigm inspired by Rommers et al. (2013), we are able to gain information on the competition process between the individual words and the idiom as a whole when the idiom-final word has yet to be integrated in the sentence context. In addition, this allows us to compare the results obtained in the current study with L2 learners to the results by native speakers presented in the previous chapter.

With respect to the semantic level of processing, in line with Cieślicka (2006), we hypothesize that the L2 learners do activate the individual word semantics, because these are more salient than the idiomatic meaning. However, as a function of increased practice and experience with the L2 idioms in the experiment, idiom processing by L2 learners may become more similar to that of native speakers. If this is the case, the idiom-final words of idioms that the L2 learners practiced intensively during the learning study, would be more and more suppressed at the semantic level. This would be in line with our findings for native speakers. However, individual target words in idioms that received only limited practice would suffer less from this suppression and would still be activated at the semantic level: For these idiomatic expressions, the meanings of the component words would be more salient than the idiom's figurative meaning.

With respect to the orthographic level of processing, we expect the idiom-final word form to be activated, because the item needs to be identified in order to verify whether the idiom is in fact presented. In Chapter 5, we reported an inhibitory effect of idiom-final noun frequency for native speakers, reflecting activation at the orthographic level and competition between the individual words and the idiom as a whole.

Based on van Ginkel and Dijkstra (2019), we predicted finding a general facilitatory effect of word frequency in the L2 learners, because the component word representations and the representation of the idiom are not strong enough to result in competition. However, this might change as a function of repeated practice and increased experience with L2 idioms, leading to an inhibitory frequency effect for idiom-final nouns that are part of idioms that received intensive practice, and a facilitatory frequency effect for idiom-final nouns that are part of idioms that received only limited practice.

## 6.2 Methods

### 6.2.1 Participants

Out of the 42 German learners of Dutch who participated in the learning experiment in Chapter 4 of this thesis, 30 took part in the current study. They were on average 20.6 years old ( $SD = 2.1$ ; Range = 19 – 29) and their proficiency level was intermediate, B1 according to CEFR. The participants had a normal or corrected-to-normal vision and were paid for participation with a gift card. This study was ethically assessed and approved by the Ethics Assessment Committee (EAC) of the Faculty of Arts of Radboud University Nijmegen (number 3382).

### 6.2.2 Materials and design

The same materials and design were used as in the experiment conducted with native speakers of Dutch described in Chapter 5. Some information about the idiom selection and target word selection was not presented in Chapter 5, because this is only relevant in the L2 context.

*Idiom selection.* The 30 idiomatic expressions used in this experiment were included in the learning experiment discussed in Chapter 4. Cross-language overlap was controlled for. In total, 7 Dutch idiomatic expressions did not have a German equivalent (NE), 18 did exist in German but in totally different content words (DW), and 5 did exist in German with some but not all content words overlapping (nW). Half of the idioms included in this experiment were practiced intensively by the German L2 learners in the learning experiment and the other half of the

idioms only received limited practice (see Chapter 4). This was counterbalanced.

*Target word selection.* As described in Chapter 5, the target words were controlled for word frequency and word length. In addition, we selected the target words in such a way that only half of them were cognates. Cognate status was computed based on the normalized Levenshtein distance as proposed by Schepens, Dijkstra, Grootjen, & van Heuven (2013). Target words with a score of 0.5 or higher were considered cognates.

See Table S5 in the Supplementary Materials for the idiomatic expressions and their corresponding target words included in the experiment.

### **6.2.3 Procedure**

Prior to this word naming experiment, the German L2 learners participated in the learning experiment presented in Chapter 4. The procedure of the word naming experiment was similar to that of Experiment 2 presented in Chapter 5. The experiment consisted of a familiarization phase and the main experiment. In the main experiment, the sentences were visually presented word-by-word. Each word was presented for 300 ms followed by a blank screen for 300 ms. Similar to Experiment 2 in Chapter 5, target word presentation was delayed and displayed after a 500 ms blank screen.

### **6.2.4 Data analysis**

First, we carried out a generalized linear mixed effects regression analysis to analyze the performance of the L2 learners in the learning experiment for the opaque idiomatic expressions included in the word naming study only, in order to assess to what extent learning took place for this specific group of idioms. Next, we performed linear mixed effects regression analyses to analyze the naming latencies obtained in the word naming experiment.

All analyses were conducted in the statistical software package 'R' version 3.4.0 (R Development Core Team, 2008), and the R packages 'lme4' (Bates et al., 2015), 'lmerTest' (Kuznetsova et al., 2017), and 'effects' (Fox, 2003) were used.

The models were built in a forward manner, starting off with a basic model including a random intercept for participants and the variables of interest (Context and Condition in the first analysis and Condition and Practice in the second analysis). Subsequently, we added different predictors to the model (random and fixed factors) one by one based on theory. After adding a predictor, we examined whether the model fit improved. If this was not the case, we decided not to include this predictor in the model. Only the final models are reported.

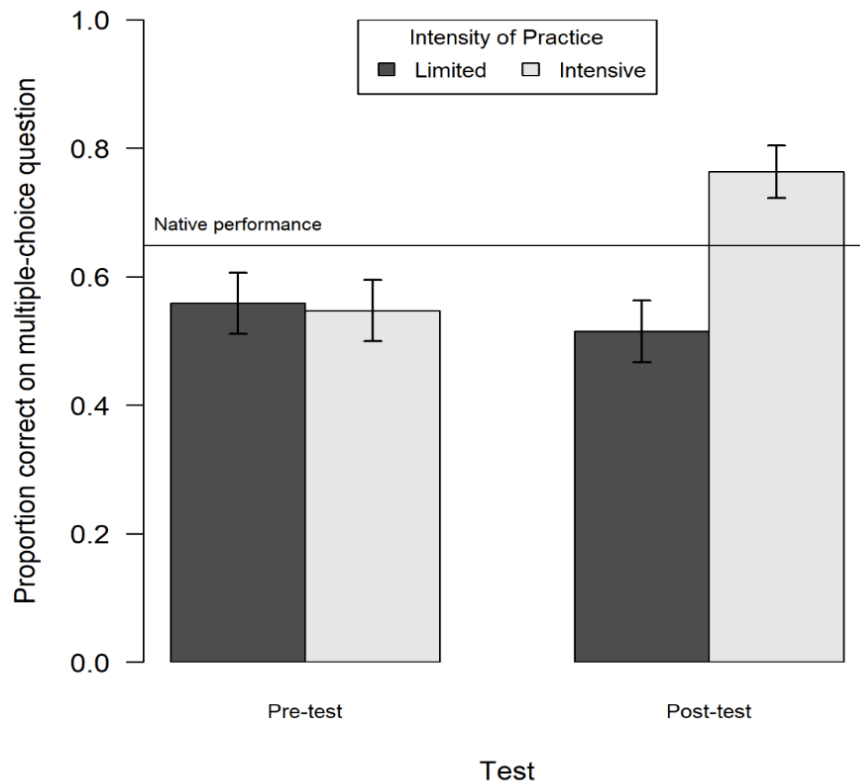
With respect to the analysis of the reaction time data, we carried out two regression analyses. One analysis was performed on the data as a whole to investigate to what extent L2 learners activate the individual words during the processing of opaque idioms in a highly biasing context. The other analyses were carried out on the idiomatic context sentences only and addressed the effect of practice (Intensive vs. Limited) on idiom processing.

## **6.3 Results**

### **6.3.1 Idiom learning**

Because only a subgroup of the L2 learners tested in the learning experiment reported in Chapter 4 participated in the word naming study, and because only the opaque idiomatic expressions from this learning experiment were included in the word naming study, the performance of this subgroup on the subset of idioms was separately analyzed. Figure 6.1 shows the average performance on the multiple-choice question targeting the meaning of the idiomatic expressions in the pre- and post-test, divided by Intensity of Practice. The horizontal line indicates the mean native performance on the same test for the same idioms (see Chapter 2).

A generalized linear mixed effects regression analysis was carried out to examine whether intensive practice with the CALL system indeed led to better idiom meaning recognition as was reported in Chapter 4 for all L2 learners participating in the learning study including all idiomatic expressions. The responses to the multiple-choice question were converted to a binary variable (correct/incorrect) and used as the dependent variable in the analysis.



*Figure 6.1* Mean proportion correct at pre-test and post-test for idioms that received limited and intensive practice. Horizontal line indicates mean native performance for the same idioms as taken from a previous study. The error bars represent SEs.

The fixed effects included in the final model were: (1) Test (reference category: Post-test), (2) Intensity of Practice (reference category: Limited), (3) Cross-Language Overlap (reference category: NE), (4) Test x Intensity of Practice, (5) Test x Cross-Language Overlap. Idioms (random intercept only) and Participants (random intercept only) were added to the model as random effects. The final model is presented in Table 6.1.

Table 6.1  
Regression Model with Idiom Knowledge as the Dependent Variable

<b>Fixed effects</b>	<b>Beta</b>	<b>Std. Error</b>	<b>z value</b>
(Intercept)	0.2339	0.2365	0.989
Pre-test	0.1650	0.2361	0.699
Intensive	1.9820	0.1579	7.586 ***
Cross-language overlap DW	-0.4469	0.2439	-1.833 .
Cross-language overlap nW	0.7806	0.3459	2.257 *
Pre-test x Intensive	-1.2476	0.2120	-5.886 ***
Pre-test x DW	0.2382	0.2545	0.936
Pre-test x nW	-0.8966	0.3589	-2.498 *
<b>Random effects</b>	<b>Variance</b>	<b>Std. Deviation</b>	
Idiom	0.1257	0.3545	
Participant	0.2555	0.5054	

Note. \* $p < 0.05$  \*\* $p < 0.01$  \*\*\* $p < 0.001$

The regression model presented in Table 6.1 showed that participants performed better at the post-test than at the pre-test on idiomatic expressions that had received intensive practice, while this was not the case for idioms that had received limited practice only ( $\beta = -1.25$ ,  $SE = 0.21$ ,  $p < 0.001$ ).

In addition, a significant effect of Test and Cross-language overlap was found, indicating that Cross-language overlap did affect performance at the post-test, but not at the pre-test. More specifically, participants performed worse on idiomatic expressions that do not exist in their L1 (NE), as opposed to idioms that have a number of content words in common (nW,  $\beta = 0.78$ ,  $SE = 0.36$ ,  $p < 0.05$ ). The difference between idioms that do not exist in the L1 (NE) and idioms that do exist in the L1, but in totally different words is only marginally significant (DW,  $\beta = -0.45$ ,  $SE = 0.24$ ,  $p < 0.1$ ).

### 6.3.2 Idiom processing

Naming errors and trials with naming latencies shorter than 300 ms and longer than 1250 ms were removed from the data, because this indicated an incorrect trigger of the voice key (13.0 %). Four participants were removed because of poor performance on the comprehension questions (< 60% correct). Responses at 2.5 SDs from the mean were removed on the participant and item level. The average naming latencies and SDs per Context and Condition are presented in Table 6.2.

Table 6.2  
Mean Naming Latencies and SDs.

Condition	Context	
	Literal	Idiomatic
COR	626 (109)	599 (113)
REL	658 (111)	645 (110)
UNREL	659 (110)	649 (102)

#### *Complete dataset*

The first linear mixed effects regression analysis was performed on the whole dataset. The dependent variable of the analysis were the log-transformed reaction times. Our final regression analysis included the



following fixed effects: (1) Context (Idiomatic and Literal), (2) Condition (COR, REL, and UNREL), (3) Trial number, (4) Initial sound (Vowels, Plosives, Fricatives, Approximants, and Nasals), (5) Target word frequency, (6) Target word length, (7) Idiom imageability, (8) Context x Condition, (9) Context x Idiom imageability. Participants (random intercept, and random slope of Trial number), and target word (random intercept only) were included as random effects. Reference categories for the categorical predictors were Literal, Vowels, and REL for Context, Initial sound and Condition respectively. The variable Idiom imageability was mean centered in order to be able to interpret the simple effect of Context. The model is presented in Table 6.3.

This regression analysis revealed that the participants responded significantly faster to the correct target words as opposed to the semantically related target words in both the idiomatic context ( $\beta = -0.083$ ,  $SE = 0.011$ ,  $p < .001$ ) and the literal context (relevelled version of the model:  $\beta = -0.046$ ,  $SE = 0.011$ ,  $p < .001$ ). However, the significant interaction effect between Context and Condition shows that the difference between the correct target words and the semantically related target words was significantly larger in the idiomatic context as opposed to the literal context ( $\beta = 0.037$ ,  $SE = 0.011$ ,  $p < .001$ ). Naming latencies were similar for semantically related and unrelated target words in both the idiomatic ( $\beta = 0.004$ ,  $SE = 0.011$ ,  $p > .05$ ) and the literal context (relevelled version of the model:  $\beta = -0.002$ ,  $SE = 0.011$ ,  $p > .05$ ). In addition, Target word frequency was observed to facilitate word naming ( $\beta = -0.01$ ,  $SE = 0.003$ ,  $p < .001$ ). Idiom imageability significantly affected the naming latencies. Although the interaction effect between Idiom imageability and Context did not reach significance ( $\beta = 0.006$ ,  $SE = 0.006$ ,  $p > .05$ ), the effect of idiom imageability was especially strong in the idiomatic context ( $\beta = -0.018$ ,  $SE = 0.007$ ,  $p < .01$ ), while it was only marginally significant in the literal context (relevelled version of the model:  $\beta = -0.012$ ,  $SE = 0.006$ ,  $p = .06$ ). Significant covariates were Target word length and Initial Sound. The longer the target words, the longer the naming latencies ( $\beta = 0.016$ ,  $SE = 0.003$ ,  $p < .001$ ), and target words starting with a vowel were named more slowly than target words starting with an approximant ( $\beta = -0.049$ ,  $SE = 0.019$ ,  $p < .05$ ) and a nasal ( $\beta = -0.055$ ,  $SE = 0.019$ ,  $p < .01$ ).

Table 6.3 Regression Model Complete Dataset with Logged Naming Latencies as the Dependent Variable

<b>Fixed effects</b>	<b>Beta</b>	<b>Std. Error</b>	<b>t value</b>
(Intercept)	6.4057	0.0279	229.932 ***
Trial number	-0.0039	0.0034	-1.172
Initial sound (Fricatives)	-0.0282	0.0161	-1.748
Initial sound (Plosives)	-0.0155	0.0157	-0.989
Initial sound (Nasals)	-0.0550	0.0193	-2.846 **
Initial sound (Approximants)	-0.0496	0.0196	-2.526 *
Target word length	0.0164	0.0032	5.16 ***
Target word frequency	-0.0103	0.0029	-3.589 ***
Context (Literal)	0.0179	0.0075	2.391 *
Condition (COR)	-0.0832	0.0111	-7.485 ***
Condition (UNREL)	0.0036	0.0111	0.326
Idiom imageability	-0.0183	0.0065	-2.828 **
Context (Literal) x Condition (COR)	0.0367	0.0105	3.484 ***
Context (Literal) x Condition (UNREL)	-0.0058	0.0106	-0.55
Context (Literal) x Idiom imageability	0.0059	0.0061	0.969
<b>Random effects</b>	<b>Variance</b>	<b>Std. Deviation</b>	<b>Corr</b>
Target word	0.0010	0.0311	
Participant	0.0091	0.0955	
Trial number	0.0002	0.0129	0.19
Residual	0.0186	0.1363	

Note. \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

*Idiomatic context only*

The second analysis investigated the effect of practice on idiom processing. Because the effect of practice was only relevant for the sentences in the idiomatic context, we performed a separate linear mixed effects regression analysis to analyze the idiomatic context sentences only. The dependent variable of the analysis were the log-transformed reaction times. Our final model included the following variables as fixed effects: (1) Condition (COR, REL, and UNREL), (2) Intensity of Practice (Intensive and Limited), (3) Trial number, (4) Initial sound (Vowels, Plosives, Fricatives, Approximants, and Nasals), (6) Target word frequency, (7) Target word length, (8) Idiom imageability, (9) Condition x Practice. As random effects we included Participants (random intercept, and random slope of Trial number) and target word (random intercept only). Reference categories for the categorical predictors were COR (for Condition), Limited (for Intensity of Practice), and Vowels (for Initial sound). The model is presented in Table 6.4.

Although including the interaction effect between Intensity of Practice and Condition only resulted in a marginally significant improvement of the model fit ( $\chi^2(2) = 5.103, p = 0.077$ ), we decided to present it here, because of its theoretical relevance. The final model, as presented in Table 6.4, showed the same pattern of results as the model based on the complete dataset with respect to the variable Condition. Participants responded faster to the correct target word as compared to the semantically related ( $\beta = 0.067, SE = 0.014, p < .001$ ) and unrelated target word ( $\beta = 0.070, SE = 0.013, p < .001$ ). The marginally significant interaction effect between Practice and Condition indicated that the effect of Practice was different for the correct target words as compared to the semantically related ( $\beta = 0.030, SE = 0.015, p < .05$ ) and unrelated target words ( $\beta = 0.029, SE = 0.015, p = .054$ ). More specifically, the correct target words that were part of idiomatic expressions that had been practiced intensively by the participants in the learning experiment (see Chapter 4) were named faster than the correct target words that were part of idioms that had received limited practice only ( $\beta = -0.026, SE = 0.011, p < .05$ ). This effect is visualized in Figure 6.2. Target word frequency and Idiom imageability had a facilitatory effect on idiom processing (resp.  $\beta = -0.012, SE = 0.003, p < .001$  and  $\beta = -0.019, SE =$

0.007,  $p < .01$ ). Similar effects were found for the covariates that significantly affected naming latencies in the analysis based on the complete dataset. Longer words were associated with longer naming latencies ( $\beta = 0.018$ ,  $SE = 0.004$ ,  $p < .001$ ), and target words starting with a vowel were responded to more slowly than target words starting with an approximant ( $\beta = -0.051$ ,  $SE = 0.023$ ,  $p < .05$ ) or a nasal ( $\beta = -0.072$ ,  $SE = 0.022$ ,  $p < .01$ ).

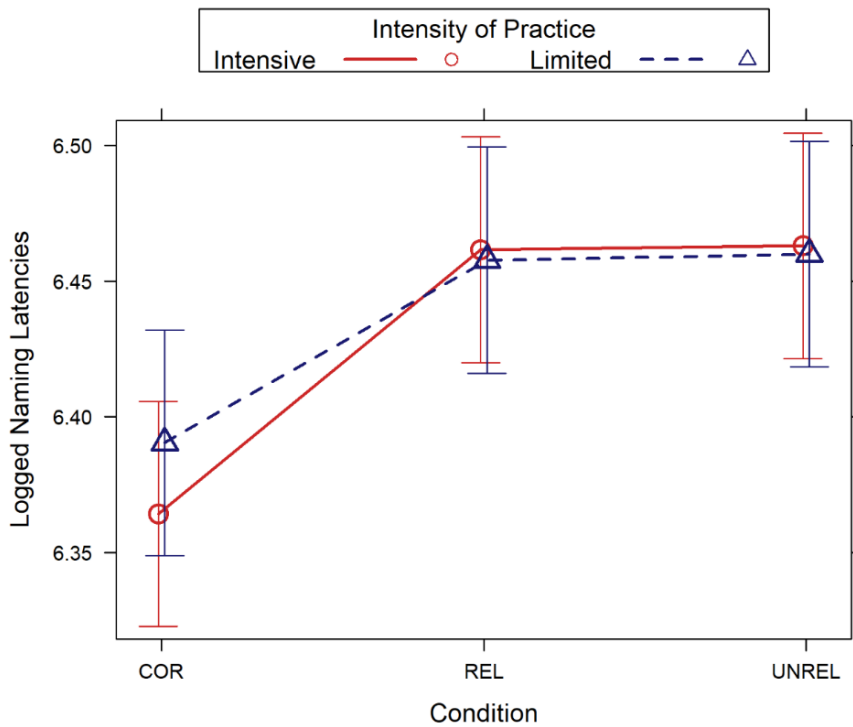


Figure 6.2 Interaction effect between Intensity of Practice and Condition. The error bars represent SEs.

Table 6.4 Regression Model Idiomatic Sentences only with Logged Naming Latencies as the Dependent Variable

<b>Fixed effects</b>	<b>Beta</b>	<b>Std. Error</b>	<b>t value</b>
(Intercept)	6.3337	0.0303	209.143 ***
Trial number	-0.0106	0.0043	-2.472 *
Initial sound (Fricatives)	-0.0276	0.0185	-1.493
Initial sound (Plosives)	-0.0219	0.0180	-1.214
Initial sound (Nasals)	-0.0721	0.0222	-3.24 **
Initial sound (Approximants)	-0.0505	0.0225	-2.242 *
Target word length	0.0177	0.0037	4.811 ***
Target word frequency	-0.0116	0.0033	-3.494 ***
Intensity of Practice (Intensive)	-0.0262	0.0107	-2.461 *
Condition (REL)	0.0673	0.0137	4.932 ***
Condition (UNREL)	0.0695	0.0135	5.163 ***
Idiom imageability	-0.0185	0.0066	-2.825 **
Intensity of Practice (Intensive) x Condition (REL)	0.0301	0.0152	1.974 *
Intensity of Practice (Intensive) x Condition (UNREL)	0.0293	0.0152	1.931 .
<b>Random effects</b>	<b>Variance</b>	<b>Std. Deviation</b>	<b>Corr</b>
Target word	0.0009	0.0315	
Participant	0.0093	0.0964	
Residual	0.0002	0.0150	0.33
	0.0186	0.1365	

Note. \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

## 6.4 Discussion

This study investigated the effects of individual word activation and exposure on the processing of opaque idiomatic expressions by L2 learners. We tested German L2 learners of Dutch using a primed word-naming task, after they had participated in a CALL-based learning study in which we systematically varied the intensity of practice of the idiomatic expressions under study (see Chapter 4). The word-naming experiment contained the same opaque idiomatic expressions that were included in the learning study. In this way, we were able to examine the activation of the component words during the processing of L2 idioms that had received either intensive practice (12 presentations) or limited practice (2 presentations) in the learning study.

### 6.4.1 Idiom learning

The L2 learners were able to learn the opaque idiomatic expressions by using the CALL system focusing on both form and meaning aspects of the idioms under study. In the post-test, participants performed significantly better than in the pre-test with respect to the idiom meaning recognition exercises for idioms that had received intensive practice. The L2 learners even outperformed the Dutch native speakers tested in our benchmark rating study on this specific set of idiomatic expressions (see Chapter 2; Hubers, Cucchiarini, Strik, & Dijkstra, 2019; Hubers, van Ginkel, Cucchiarini, Strik, & Dijkstra, 2018). Performance on the idioms that were practiced only a few times did not improve.

In addition, effects of cross-language idiom overlap were found. The L2 learners performed better on Dutch L2 idiomatic expressions that had exact equivalents in form and meaning in German (i.e., Dutch: *de eerste viool spelen*, German: *die erste Geige spielen*) than on idioms that did not exist in German (neither form nor meaning overlap) (i.e., Dutch: *goed uit de verf komen*, German: no equivalent), or that did exist in German, but consisted of totally different content words (no form overlap) (Dutch: *hoog van de toren blazen*, German: *große Töne spucken*). However, this effect was only visible in the post-test. This is in line with our results in Chapter 4, and indicates that the participants initially did

not exploit their L1 knowledge, but learned to use this information as a result of practice with the idioms.

#### **6.4.2 Idiom processing**

Idiomatically biasing contexts led to faster naming latencies for the correct target word (the idiom-final noun) than the semantically related and unrelated target words. This finding implies that L2 learners recognized the idiomatic expressions as such, and is in support of a separate idiom representation. In the literal context, a similar facilitation effect was found for the correct target word, which shows that the L2 learners use the sentence context to anticipate upcoming words. These findings are in line with findings for native speakers using the same paradigm (see Chapter 5, and Rommers et al., 2013). Moreover, they agree with Beck and Weber (2016), and van Ginkel and Dijkstra (2019), who showed that L2 learners in cross-modal priming tasks have access to the figurative meaning as reflected by a priming effect of target words that are related to the idiom meaning.

Furthermore, we found that the semantically related and unrelated target words did not differ in terms of naming latencies in the idiomatic context. This suggests that L2 learners did not substantially activate the semantics of the idiom-final word during the processing of opaque idiomatic expressions. This is in contrast with the Literal Salience Model by Cieślicka (2006), which assumes primacy of the literal word meanings over the figurative meaning. However, we are unable to interpret the lack of this effect unequivocally, because a similar pattern of results was observed in the literal context. In this context, the activation of the correct target word was expected to spread to semantically related target words, but this did not occur.

These results are similar to those obtained for native speakers in Experiment 1 discussed in Chapter 5. For native speakers, we reasoned that activation may not have had enough time to spread from the correct target word to the semantically related words. Therefore, we delayed the target word presentation by 200 ms in a second experiment, which led to the expected graded pattern of results in the literal context (COR < REL < UNREL). For the experiment with L2 learners, we applied the same target word presentation times as in the second experiment with native

speakers, but in this case a different picture emerged in the literal context. One potential explanation is that the delay of 200 ms was not enough to increase the pre-activation of the correct target word in L2 learners, possibly because second language processing generally takes more time than L1 processing (e.g., Gollan et al., 2011; Lemhöfer et al., 2008). Another possibility is that the individual word representations were not strong enough for the activation to spread to semantically related words.

Although these data make it difficult to ascertain whether the idiom-final noun is semantically activated or not, they do show that L2 learners did activate the idiom-final noun orthographically. The observed effect of target word frequency testifies to this. The target word has to be identified in order to verify whether the incoming sequence is an idiom or a string of words without a figurative meaning. The ease with which this is done depends on the idiom's frequency of usage. A facilitatory effect of word frequency was observed for L2 learners, suggesting that an idiom-final noun with a higher word frequency is easier to integrate in the idiomatic context than a low frequency idiom-final noun.

These results are in line with our hypothesis and suggest that at the orthographic level, the idiom-final noun (the correct target word in the idiomatic context) is activated. Based on these results L2 learners can be likened to native speakers. However, L2 learners differed from the native speakers with respect to the *direction* of the word frequency effect. While L2 learners showed a facilitatory effect of idiom-final noun frequency, native speakers displayed an inhibitory effect of idiom-final noun frequency (Chapter 5). A similar difference between L2 learners and native speakers has been reported by van Ginkel and Dijkstra (2019). They argued that the absence of a competition effect in L2 learners should be ascribed to their generally lower degree of exposure to the L2, which in turn leads to weaker representations of the individual words and the idiom as a whole.

While intensive practice with L2 idioms led to increased performance on the off-line idiom recognition test and L2 learners even outperformed native speakers on this task, the amount of practice did not significantly affect the processing of these idiomatic expressions by L2 learners. Repeated practice did not strengthen the L2 idiom's representation and the representations of the component words in such



a way that at the orthographic level competition could be observed. However, a trend was visible in response to the idiom-final nouns in the idiomatic context: The idiom-final nouns that were part of idioms that had been practiced intensively were responded to faster than the idiom-final nouns of idioms that had received limited practice. This suggests that repeated practice facilitates accessing the representation of the idiom meaning. Future research might confirm this finding.

Another important difference between the learning and the word naming results is that the degree of cross-language overlap had an impact on off-line idiom comprehension, but did not affect on-line L2 idiom processing.

This finding is in contrast with Titone, Columbus, Whitford, Mercier, and Libben (2015) who reported an effect of cross-language overlap on L2 idiom processing. In their preliminary experiment, English-French bilinguals read English sentences in an RSVP-paradigm. On the last word of the sentence, participants had to decide whether the sentence was meaningful or not. The sentences either contained English idioms or were matched literal controls. The sentence-final word was presented in English or was translated into French to create a code-switched version of each sentence (e.g., *He played with feu/fire*). Titone et al. (2015) found a positive effect of cross-language overlap in the code-switched condition. Participants responded faster to the meaningfulness question if the degree of cross-language overlap was higher. In the non-code-switched condition (the sentence-final word was presented in English), the effect of cross-language overlap was not significant, although a trend was observed in the same direction as in the code-switched condition.

According to Titone et al. (2015), the experiment possibly was not sensitive enough to detect an effect of cross-language overlap in the non-code-switched condition. Note that mainly bilinguals were tested with English as their L1. Thus, cross-language overlap effects may have been suppressed, given that the English idioms were highly familiar to the participants. This explanation, however, seems unlikely to hold for our findings, because the L2 learners were native speakers of German and were not as highly proficient in their L2 as the English-French bilinguals.

Another explanation by Titone et al. (2015) pointed to the nature of the judgement decision. This meaningfulness decision only required a superficial response. Participants were not required to interpret the idioms in any specific way. In our study, we used a different task and did not find effects of cross-language overlap either. This word-naming task, however, did not require the participants to make a decision about the idiom either. Possibly, an effect of cross-language overlap might arise in an idiom-decision task, for example, the idiom-equivalent of a lexical-decision task, or if we translated the target words into German, the L1 of our participants. More research is necessary to study the effect of cross-language overlap on L2 idiom processing in more detail.

From the results of the word naming experiment, it is not yet clear to what extent idiom-final nouns are activated at the semantic level of representation during the processing of opaque idiomatic expressions. At the orthographic level of representation, however, we found that idiom-final nouns are activated, as indicated by the facilitatory effect of word frequency. Just like native speakers, L2 learners must identify the orthography of the idiom-final word to verify that an idiom is present. However, this effect is different from the competition effect observed in native speakers, and did not change through repeated exposure.

The results suggest that L2 learners are able to access the figurative meaning during the processing of opaque idiomatic expressions. How quickly this is done may be affected by exposure to or practice with L2 idiomatic expressions.

## **6.5 Conclusions**

The present study investigated the role of the individual words during L2 idiom processing and the effect of exposure to or practice with L2 idiomatic expression on L2 idiom processing. To this end, techniques from both applied linguistics and psycholinguistics were combined. L2 idiom processing was investigated in German L2 learners of Dutch after they participated in a CALL-based experiment aimed at increasing their experience with L2 idioms. We manipulated intensity of practice to be able to determine to what extent L2 idiom processing is affected by exposure.

The study reported on in this Chapter leads to useful insights about L2 idiom learning and L2 idiom processing. First, from a theoretical point of view, the study indicates that native speakers and L2 learners may employ similar underlying mechanisms in processing L2 idiomatic expressions. After sufficient exposure, a similar performance can be observed in native speakers and L2 learners. In the off-line idiom knowledge test L2 learners even outperform native speakers after intensive practice. Second, the word naming experiment also reveals *differences* in performance between native speakers and L2 learners. Native speakers and L2 learners show a different effect of word frequency. While native speakers display an inhibitory frequency effect on the idiom-final noun, reflecting a competition process between the individual words and the idiom as a whole, L2 learners show a facilitatory frequency effect on the idiom-final noun.

While intensive practice did not lead to similar results for L2 learners as compared to native speakers in terms of on-line idiom processing, the amount of practice was sufficient for L2 learners to even outperform native speakers in the off-line idiom knowledge test.



## Chapter 7

# **General Discussion and Conclusions**

This thesis intends to increase our understanding of the knowledge, representation, learning, and processing of idiomatic expressions by L2 learners relative to native speakers. More specifically, we considered to what extent L2 learners (1) are comparable to native speakers in terms of idiom knowledge, representation, and processing, (2) can become more similar to native speakers as a function of practice, and (3) are sensitive to properties of idioms in comparison to native speakers.

In Section 7.1, we will answer these questions and discuss the thesis results in light of a comprehensive model of L1 and L2 idiom processing. Next, we discuss the significance of the research apparatus developed in this thesis (Section 7.2), the relevance of this thesis for scientific research (Section 7.3) and teaching practice (Section 7.4), and present methodological recommendations based on the thesis results (Section 7.5). Next, we describe perspectives for future research (Section 7.6) followed by conclusions in Section 7.7.

## **7.1 General Discussion**

In the current section we will first answer the research questions based on results described in the various Chapters of this thesis (Sections 7.1.1-3). Subsequently, we integrate the answers to these questions in a theoretical framework. To this end, we first extend the monolingual hybrid idiom processing model by Sprenger et al. (2006) to L2 learner processing (Section 7.1.4). Next, we interpret the various empirical studies we conducted within the framework of the new model, taking into account the role of experience, idiom properties, and the individual words (subsections 7.1.4).

### **7.1.1 To what extent are L2 learners comparable to native speakers in terms of knowledge, representation, and processing of idiomatic expressions?**

To properly assess the L2 learners' knowledge and use of idioms, we first investigated idiom knowledge in native speakers. In the comprehensive study in *Chapter 2*, native speakers rated Dutch idiomatic expressions on various important idiom properties, such as familiarity, transparency, and imageability. The study also objectively assessed their receptive

knowledge of the idiom meanings in a multiple-choice question test (meaning recognition). The receptive idiom knowledge of native speakers was considerable, given that they were generally able to correctly identify the meaning of the idiomatic expressions, and their intuitions about the experience-based dimensions (familiarity, frequency, and usage) were highly developed. In **Chapter 3**, we investigated the same aspects in L2 learners of Dutch and compared them to native speakers. Receptive idiom knowledge appeared to be much lower for L2 learners than for native speakers. Moreover, we observed large differences between the intuitions of native speakers and L2 learners with respect to the experience-based dimensions familiarity, frequency, and usage. These findings confirm the general belief that L2 learners have less experience with idiomatic expressions than native speakers.

With respect to idiom processing, we next investigated the role of the individual words during the online processing of opaque idiomatic expressions in native speakers and L2 learners. In a primed word-naming task (reported in **Chapter 5**), we found that in native speakers the component words of opaque idiomatic expressions are in competition with the idiom as a whole on different levels of representation. At the semantic level, the individual words were not activated. Facilitatory effects were probably cancelled out due to suppression of the individual word meanings. The individual word meanings appeared to be suppressed, possibly because it would otherwise interfere with the meaning of the idiom as a whole. However, activation of the individual words did persist at the orthographic level, because to verify whether an idiom is actually being presented, the words constituting the idiom must be identified. In sum, Chapter 5 provides evidence in favor of a hybrid model of idiom processing in which the individual words and the idiom as a whole interact at both the orthographic and semantic levels of representation.

In **Chapter 6**, we looked at the role the individual words in idioms play for L2 learners, carrying out the same experiment as in Chapter 5 for the native speakers. The results did not allow us to conclude that individual words are activated semantically in the L2 learners during the processing of opaque idiomatic expressions. But L2 learners, like native

speakers, did activate the individual words' orthographic forms, as shown by a word frequency effect. However, the direction of the word frequency effect was different for the two participant groups: Whereas L2 learners showed a facilitatory effect of word frequency, native speakers showed an inhibitory effect. This inhibitory effect suggests the presence of a competition process between the individual words and the idiom as a whole in native speakers. The absence of this effect in L2 learners may be due to their limited experience with the idiomatic expressions, possibly insufficient to trigger a competition process.

To summarize, several results in this thesis point to different outcomes for L2 learners as compared to native speakers in terms of both idiom knowledge and idiom processing. However, the observed differences might gradually disappear with more experience with the L2 in general and L2 idioms in particular. If this is indeed the case, this suggests that native speakers and L2 learners use the same underlying architecture to process idiomatic expressions. Differences between L2 learners and native speakers can then be explained in terms of differences in experience. This issue will be considered next.

### **7.1.2 To what extent can L2 learners become more similar to native speakers as a function of practice?**

On the basis of some of the results in this thesis, we would like to argue that, although L2 learners differ from native speakers with respect to idiom knowledge and idiom processing, they can indeed become more similar to native speakers as a function of practice.

In *Chapter 4*, we investigated the effect of practice intensity on receptive L2 idiom knowledge. We adopted a new experimental paradigm in which German L2 learners used a Computer Assisted Language Learning (CALL) system providing practice on form and meaning aspects of Dutch idiomatic expressions and immediate corrective feedback. Before the training, a pre-test was administered in which their receptive knowledge of Dutch idiomatic expressions was assessed using the same procedure as in Chapters 2 and 3. During training, we manipulated intensity of practice, such that half of the idioms received intensive practice (12 presentations), while the other half received limited practice (2 presentations). After training, a post-test was



administered testing the participants' idiom knowledge again. We found that intensive practice enhanced L2 idiom knowledge. L2 learners could even achieve native-like performance. Even limited practice involving only two presentations of each idiom led to increased L2 idiom knowledge. These results showed that, if sufficient exposure and active practice are provided, L2 learners do manage to acquire L2 idiomatic expressions, and they are even capable of outperforming native speakers.

In *Chapter 6*, we investigated the effect of practice on the activation of individual words during the processing of opaque idiomatic expressions. The German L2 learners of Dutch who had participated in the learning study (Chapter 4) subsequently took part in a psycholinguistic experiment similar to that conducted with native speakers (Chapter 5). The results showed that individual word activation does not significantly change as a function of practice. However, we *did* find that L2 learners are able to access the idiom meaning. How quickly they do so seems to be affected by intensive practice with L2 idiomatic expressions.

Thus, these results show that differences between native speakers and L2 learners in terms of idiom knowledge and idiom processing are not insurmountable, and can be overcome if sufficient exposure to L2 idiomatic expressions is provided.

### **7.1.3 To what extent are L2 learners sensitive to the same idiom properties as native speakers?**

In spite of potential differences, various chapters in this thesis suggest that L2 learners are sensitive to the same idiom properties as native speakers. Similar effects of transparency and imageability are observed for native speakers and L2 learners. In addition, cross-language overlap is relevant for L2 learners.

*Chapters 2 and 3* showed that idiom transparency positively affected both L1 and L2 idiom knowledge, indicating that receptive idiom knowledge increased if the idiom was more transparent. This shows that both native speakers and L2 learners make use of the idiom's component words whenever possible. For native speakers, transparency was especially important if they felt less familiar with the idiom. In L2 learners, transparency was, in general, the strongest determinant of

idiom knowledge. These effects can be considered similar. Transparency becomes more important if one has less experience with the idiom. L2 learners have in general less experience with L2 idioms than native speakers, which results in a strong overall effect of transparency. This idea was confirmed in our CALL-based learning study in **Chapter 4**. After intensive practice, transparency did not affect L2 learners' receptive idiom knowledge, while after limited practice, transparency *did* turn out to be an important determinant of L2 idiom knowledge.

**Chapters 2 and 3** showed a negative effect of imageability for both native speakers and L2 learners: the better the participants could visualize the idiom, the worse their performance on the idiom knowledge test. For native speakers, this negative effect was stronger if they indicated being more familiar with the idiomatic expression. We argued that both native speakers and L2 learners probably visualize the literal interpretation of the idiom and that this is likely to interfere with correctly identifying the figurative meaning. Visualizing the literal interpretation of the idiom *to kick the bucket* may hinder idiom processing and correct recognition of the idiom meaning, for instance because one could imagine an angry farmer kicking a bucket, which is not related to the figurative meaning 'to die'. However, if one knows the etymology, visualizing the literal interpretation should facilitate performance on the idiom knowledge test, instead of hindering it. In this case, one could imagine a person standing on a bucket who is about to get hanged at the gallows. The image of someone kicking the bucket is then associated to the figurative meaning 'to die'. This shows that visualizing the literal interpretation does not necessarily hinder idiom processing and performance on the idiom knowledge test, but could work as a mnemonic device. This makes it difficult to interpret the negative effect of imageability in relation to the literal meaning. It is clear that more focused research is needed in order to find out what it is that participants visualize in the case of idiomatic expressions and how these visualizations affect idiom knowledge.

With respect to idiom processing, we found that both native speakers and L2 learners are sensitive to imageability (**Chapters 5 and 6**). While we observed a negative effect of imageability on L1 and L2 idiom knowledge, imageability facilitated the processing of opaque

idiomatic expressions. This facilitatory effect can be due to the fact that we only included opaque idiomatic expressions in our processing experiments. For these expressions, participants may have visualized the figurative meaning more often than the literal meaning, leading to facilitated idiom processing.

A factor that is especially relevant for L2 learners is cross-language overlap. The degree to which an L2 idiomatic expression has an equivalent in the L1 affected L2 idiom knowledge, but did not affect L2 idiom processing. In *Chapter 3*, we found that only L2 idioms that have a word-to-word correspondent in the L1 are associated with higher scores on the idiom knowledge test. *Chapter 4*, however, revealed that after training, L2 learners also show an advantage in terms of receptive idiom knowledge for L2 idioms that have an L1 equivalent that is very similar, but not identical in form. This suggests that during training, L2 learners become aware of the similarities between the L1 and the L2 in the case of languages that are relatively closely related (Dutch and German).

In terms of idiom processing, no effect of cross-language overlap was found in L2 learners (*Chapter 6*). This suggests that during L2 idiom processing, L2 learners do not have time to apply their knowledge about L1 idioms to on-going processing.

To summarize, native speakers and L2 learners are sensitive to both idiom transparency and imageability. The extent to which these idiom properties are used to access the figurative meaning seems to depend on the participants' experience with the idioms. Moreover, L1 knowledge, as reflected by cross-language overlap, is an important source of information for L2 learners, who benefit from it if the idiomatic expressions in the L1 and L2 are similar both in meaning and form.

#### **7.1.4 A comprehensive model of L1 and L2 idiom processing**

In contrast to researchers who argued that L2 learners are fundamentally different from native speakers when it comes to idiom processing (Abel, 2003; Cieślicka, 2006; Wray, 2002), the results presented in this thesis suggest that native speakers and L2 learners make use of the same underlying architecture to process idiomatic expressions. Both native speakers and L2 learners are sensitive to exposure and experience, and idiom properties such as transparency and imageability similarly affect

idiom processing in native speakers and L2 learners. The differences between native speakers and L2 learners observed with respect to idiom knowledge, representation, and processing seem to derive from differences in degree of experience with L2 idioms. The L2 learners' more limited exposure and experience with the L2 in general and L2 idioms in particular would be responsible for their more limited knowledge, generally slower processing of L2 idioms, and for their slightly less reliable intuitions about idiom properties.

This suggests that one comprehensive model can be proposed for both L1 and L2 idiom processing, in which processing speed and strength of representations explain the observed differences between native speakers and L2 learners. In this section, we discuss such a model. Although other idiom processing models might also account for the data obtained in this thesis (Abel, 2003; Cacciari & Tabossi, 1988; Carrol & Conklin, 2017), we have chosen to adopt the existing hybrid idiom production model for native speakers by Sprenger et al. (2006) and to extend it to idiom comprehension in L2 learners. Before introducing the bilingual model, we first briefly describe the model by Sprenger et al. (2006).

The hybrid monolingual model of idiom processing as proposed by Sprenger et al (2006) is applied in Figure 7.1 to the Dutch idiomatic expression *het loodje leggen* (lit. 'to lay the piece of lead') meaning 'to die'. In this model, the idiom has a separate entry in the mental lexicon (the superlemma). This idiom representation is connected to its idiom component word lemmas on the one hand (the simple lemmas), and its meaning representation on the other hand. The model allows for parallel activation of the individual words and the idiom as a whole. In addition, activation of the simple lemmas *loodje* ('piece of lead') and *leggen* ('to lay') can activate the superlemma *het loodje leggen* ('to die'), and the other way around. At the conceptual level, activation can spread from one node to semantically related concepts. Recent studies investigating idiom processing have obtained results in support of such a model in native speakers (e.g., Libben and Titone, 2008; Sprenger et al., 2006; Titone and Libben, 2014).

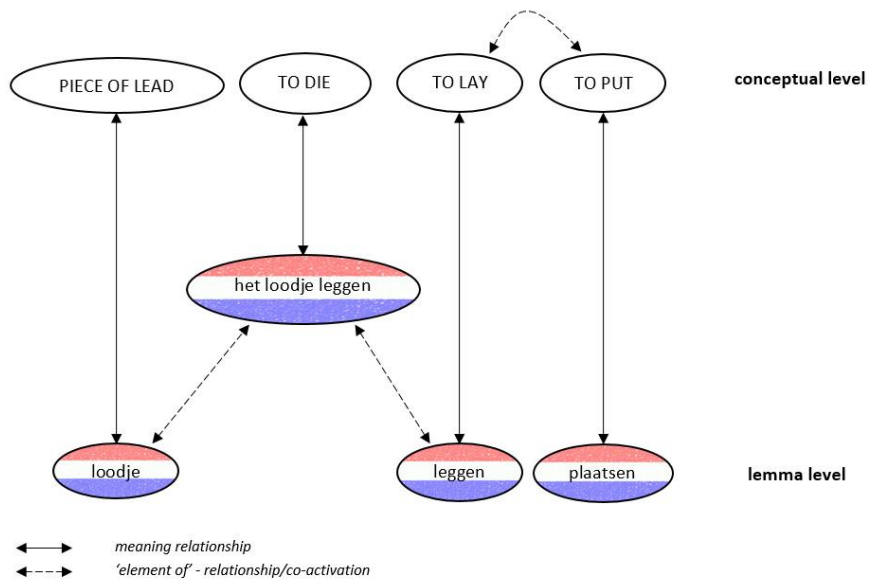


Figure 7.1. The monolingual model by Sprenger et al. (2006), applied to the Dutch idiom *het loodje leggen* ('to die').

We propose a bilingual version of the model that can account for idiom processing in both native speakers and L2 learners. Figure 7.2 shows an instantiation of this model for the Dutch idiom *het loodje leggen* and its connections in an English L2 learner of Dutch. As in the model by Sprenger et al. (2006), the idiom has a separate representation at the lemma level (the superlemma) that is connected to its figurative meaning representation. In addition, the simple lemmas are connected to the superlemma allowing for co-activation.

The monolingual version of the model in Figure 7.1 is incorporated in the bilingual version (see the left side of Figure 7.2). In addition to the L2 idiom representation, the model also includes an L1 idiom equivalent if available. The L1 and L2 superlemmas are both connected to the same figurative meaning representation.

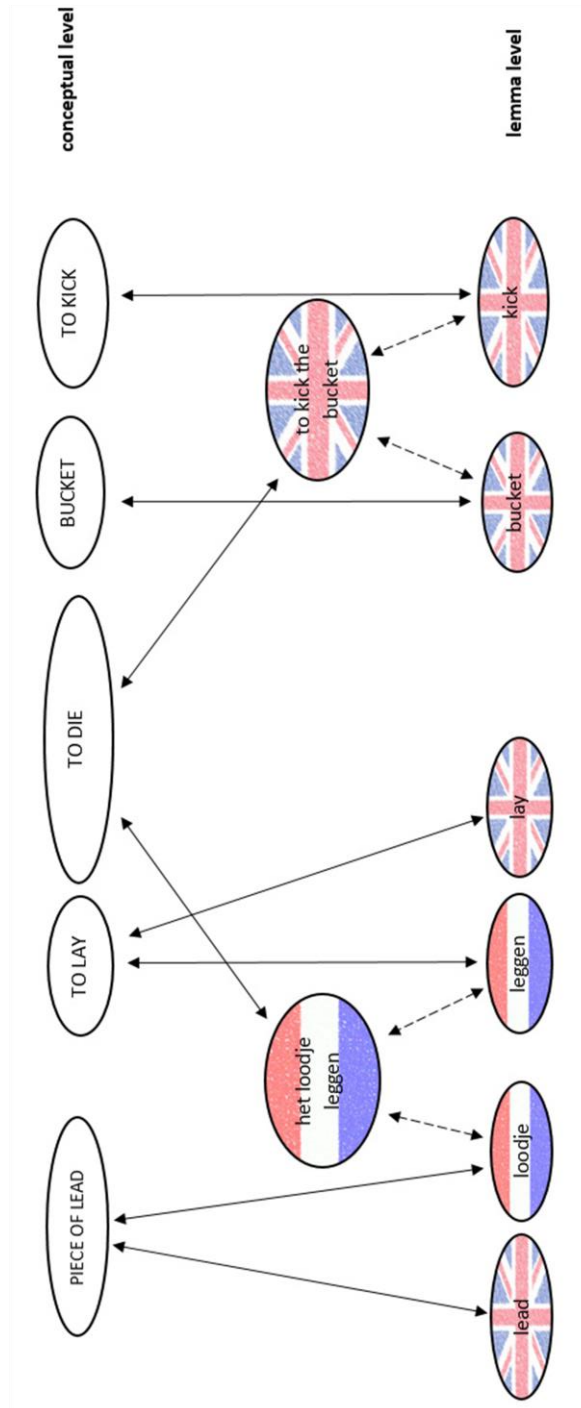


Figure 7.2 Bilingual version of the idiom processing model by Sprenger et al. (2006)

In line with Sprenger et al. (2006), the model only represents the lemma level and the conceptual level. However, the simple lemmas are also connected to their corresponding orthographic representations. We assume a more or less direct relationship between the orthography of the individual words and their word lemmas. Moreover, after activating the meaning representations, they are integrated into an event structure (cf. Levelt, 1989) that, in turn, is incorporated in a mental model. It is on the level of the event structure that the literal meaning of the idiom as a whole is computed. We consider lemmas to be language specific (de Bot, 1992; de Bot, Cox, Ralston, Schaufeli, & Weltens, 1995; Wei, 2002). Just as in the monolingual model by Sprenger et al. (2006), the superlemmas contain information about their syntactic structure and potentially modify the syntactic features of the simple lemmas they consist of.

In Figure 7.2, we presented only one instantiation of the bilingual idiom processing model. However, the thesis results suggest that model variants can take into account the amount of experience with the idiom, the specific properties of the idiom (i.e., transparency and cross-language overlap) and aspects of the idiom's component words. In the following subsections, we describe how the results lend support to the proposed bilingual idiom processing model by considering these different aspects, for each of which thesis results will be discussed in relation to both native speakers and L2 learners.

#### **7.1.4.1 The role of experience**

The amount of experience with the idiom may affect the strength of the idiom representation and its connections to the idiom meaning on the one hand and the individual words on the other. This idea was already put forward by Abel (2003) in her Dual Idiom Representation Model aimed at L2 learners. She argued that L2 learners can develop separate representations for the L2 idioms if the idioms are encountered frequently enough. In her model, differences between native speakers and L2 learners with respect to idiom knowledge are mainly explained in terms of the familiarity with L2 idioms: Native speakers have developed more separate idiom representations because they have been more frequently exposed to these expressions than L2 learners.

In the current bilingual idiom processing model, we have incorporated this idea. The thesis results suggest that the amount of experience with

the idiom affects the idiom representation in both native speakers and L2 learners.

#### *Native speakers*

Although native speakers generally have ample experience with idiomatic expressions, they still vary in terms of their intuitions and receptive idiom knowledge (**Chapters 2 and 3**). The intuitions about the experience-based idiom properties (familiarity, frequency, and usage) seem to reflect the strength of the idiom representation. Low scores on these idiom properties can be associated with weaker idiom representations. Moreover, a weaker idiom representation is associated with less detailed knowledge of the idiom's meaning. The strength of the connections between the superlemma of the Dutch idiom *het loodje leggen* and its corresponding meaning representation in Figure 7.2 may depend on the amount of experience a Dutch native speaker has with the idiom. In addition, this also seems to be the case for the connections between the simple lemmas of which the idiom consists and the superlemma.

Thus, relatively limited experience with idioms results in weaker idiom representations and weaker connections between the simple lemmas and the superlemma, and between the superlemma and the idiom meaning representation, while more experience leads to stronger idiom representations and connections.

#### *L2 learners*

The L2 idiom representations in L2 learners are weaker than those in native speakers, as indicated by the L2 learners' lower scores on the experience-based idiom properties (**Chapter 3**). More specifically, the connection between the L2 superlemma and its corresponding meaning representation and the connections between the L2 simple lemmas and their corresponding superlemma are less strong for L2 learners than for native speakers. This explains the difference in performance on the idiom knowledge questions between the native speakers and the L2 learners in Chapter 3.

However, the L2 idiom representations and their connections can become stronger over time as a function of increased experience with the L2 idioms. This was shown in **Chapter 4**. Intensive practice and repeated



exposure led to a better knowledge of the idioms' meanings in L2 learners, even for the idiomatic expressions that only received limited practice.

Our research on L2 idiom processing also suggests that L2 idiom representations can become stronger as a result of increased experience with L2 idiomatic expression. L2 learners were faster to access the representations of idiomatic expressions that they had intensively practiced as opposed to the idioms that they had encountered only twice (*Chapter 6*).

Thus, L2 idiom representations and their connections are weaker in L2 learners as compared to native speakers. However, as reflected in L2 idiom knowledge and processing, these representations can become stronger after increased experience with L2 idioms.

#### **7.1.4.2 The role of idiom properties**

In addition to practice and exposure, the specific properties of individual idioms also affect the degree to which they are known and processed. Throughout this thesis, we observed effects of idiom transparency (i.e., to what extent the idiom's component words contribute to the idiomatic meaning), idiom imageability (i.e., to what extent the idiom can be visualized) and cross-language overlap (i.e., the degree of similarity between idioms in L1 and idioms in L2). Transparency and imageability turned out to be important for both native speakers and L2 learners, while cross-language overlap was relevant for L2 learners only. Because it is not completely clear what it is that participants visualize in the case of idiomatic expressions, we will not try to explain the effects of imageability in terms of the bilingual idiom processing model at this stage.

In the current section, we first describe how the effect of transparency can be accounted for by the bilingual idiom processing model in both native speakers and L2 learners. Next, we will discuss different instantiations of the model depending on the degree of cross-language overlap. Table 7.1 shows an overview of the levels of cross-language overlap we distinguished in this thesis with examples from Dutch and English.

Table 7.1  
 Overview of the categories of cross-language overlap distinguished in this thesis.

Category	Description	L2 Dutch	L1 English
AW (All Words) (Figure 7.3)	L2 idiom has a word-to-word correspondent in L1; Identical in form and meaning.	<i>de eerste viool spelen</i> (lit. 'to play the first fiddle')	<i>to play (the) first fiddle</i>
nW ( <i>n</i> Words) (Figure 7.4)	The L2 idiom does exist in L1, and has <i>n</i> words in common, but not all; Partly overlap in form, identical in meaning	<i>de koe bij de hoorns</i> <i>vatten</i> (lit. 'to grab the cow by the horns')	<i>to take the bull by the horns</i>
DW (Different Words) (Figure 7.2)	The L2 idiom does exist in L1, but in completely different content words; No overlap in form, identical in meaning	<i>het loodje leggen</i> (lit. 'to lay the piece of lead')	<i>to kick the bucket</i>
NE (No Equivalent) (Figure 7.5)	The L2 idiom does not exist in L1; No overlap in form and meaning.	<i>goed uit de verf komen</i> (lit. 'to come well out of the paint')	not available

*Native speakers*

Transparency positively affected idiom knowledge in native speakers and was especially important when familiarity with the idiom was lower (**Chapter 2**). In the bilingual idiom processing model the figurative meaning of a transparent idiom can be retrieved in two ways. The idiom meaning can be accessed (1) via its corresponding superlemma (the direct route), and (2) via its component word meanings (the indirect route). These individual word meanings can be combined in an event structure corresponding to the literal interpretation of the idiom. In the case of transparent idioms, this event structure will be closely related to that of the figurative meaning. Because of this similarity, the idiom's figurative meaning can be retrieved. The direct route is to access the figurative meaning via the superlemma. If native speakers have only little experience with the idiom, however, the idiom representation may be not strong enough to make use of this direct route. In that case, the figurative meaning has to be retrieved via the individual word meanings.

For opaque idiomatic expressions, in which the figurative meaning cannot be retrieved from the component words, the indirect route does not lead to the figurative meaning. In this case, the literal and the figurative interpretation cannot be integrated in a similar event structure, on the basis of which the figurative meaning can be extracted. The only available route to retrieve the figurative meaning is the direct route via the superlemma. If the representation of such an idiom is relatively weak, however, it is more difficult to retrieve its figurative meaning, leading to worse performance on the idiom knowledge test.

In the case of opaque idiomatic expressions the individual word meanings might even interfere with the meaning of the idiom as a whole. Evidence in favor of this interpretation comes from the idiom processing experiment presented in **Chapter 5**. During the processing of opaque idiomatic expressions the individual word meanings were found to be suppressed in native speakers. Important to note is that the idiomatic expressions were embedded in highly biasing contexts, which may have strengthened the suppression effect.

*L2 learners*

Transparency positively influenced receptive idiom knowledge in L2 learners, just as in native speakers, and was even the most important determinant of L2 idiom knowledge (**Chapter 3**). Because of their more limited experience with L2 idiomatic expressions and consequent weaker representations, L2 learners have to rely even more on the idiom's component words than native speakers. In these cases, they have to resort to the indirect route to retrieve the figurative meaning.

After practice with L2 idioms, transparency only influenced L2 idiom knowledge for idioms that had received limited practice (**Chapter 4**). In terms of the bilingual idiom processing model, intensive practice strengthened the direct route in such a way that in the case of both transparent and opaque idioms, L2 learners did not have to rely on the indirect route to access the idioms' figurative meanings. In contrast, for the expressions that had received only limited practice, L2 learners used the indirect route, via the individual word meanings, to access the idiomatic meaning. However, there is only a reasonable chance of successfully finding the idiom's meaning via this route if the idiom is transparent. Hence, an advantage for transparent over opaque idiomatic expressions in terms of idiom knowledge arose after limited practice.

The presence of L1 idiom representations in the bilingual idiom processing model may depend on the degree of cross-language overlap for specific language pairs and idiomatic expressions. Evidence in favor of this idea comes from the observation that L2 idioms that have a direct translation equivalent in the L1 (AW expressions) are better known by L2 learners than L2 idioms that do not have a word-to-word correspondent in the L1 (**Chapters 3 and 4**). The bilingual idiom processing model, therefore, is slightly different for AW, nW, DW and NE expressions in terms of concepts, idiom representations and their connections (see Table 7.1).

Figure 7.3 represents an instantiation of the model for an English L2 learner of Dutch in the case of an L2 idiomatic expression that has a word-to-word correspondent in the L1 English (an AW expression). Both the L1 and the L2 superlemma are connected to the same idiom meaning representation, i.e., 'to be very important'. In addition, both superlemmas are connected to their language specific simple lemmas. Activating the

Dutch simple lemmas *viool* ‘violin’ and *spelen* ‘to play’ will activate the superlemma *de eerste viool spelen* if the idiom representation is strong enough. If, however, the expression is not familiar to L2 learners, its figurative meaning can still be retrieved in an indirect way. The Dutch simple lemmas will activate the corresponding nodes at the conceptual level. Activating the individual word meanings may lead to activation of the corresponding English simple lemmas, i.e. *fiddle* and *play*. These English simple lemmas can, in turn, activate the English superlemma *to play (the) first fiddle*, which then leads to retrieval of the figurative meaning. Even if L2 learners have limited experience with the L2 idiomatic expression, they are able to correctly identify the idiom meaning (**Chapter 3**).

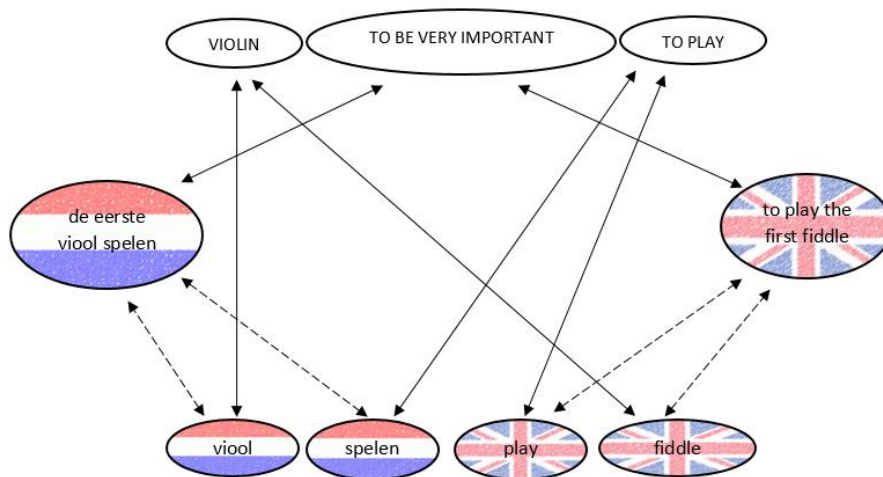


Figure 7.3 Instantiation of the model for an English L2 learner of Dutch in the case of an L2 idiomatic expression that has a word-to-word correspondent in the L1 (AW)

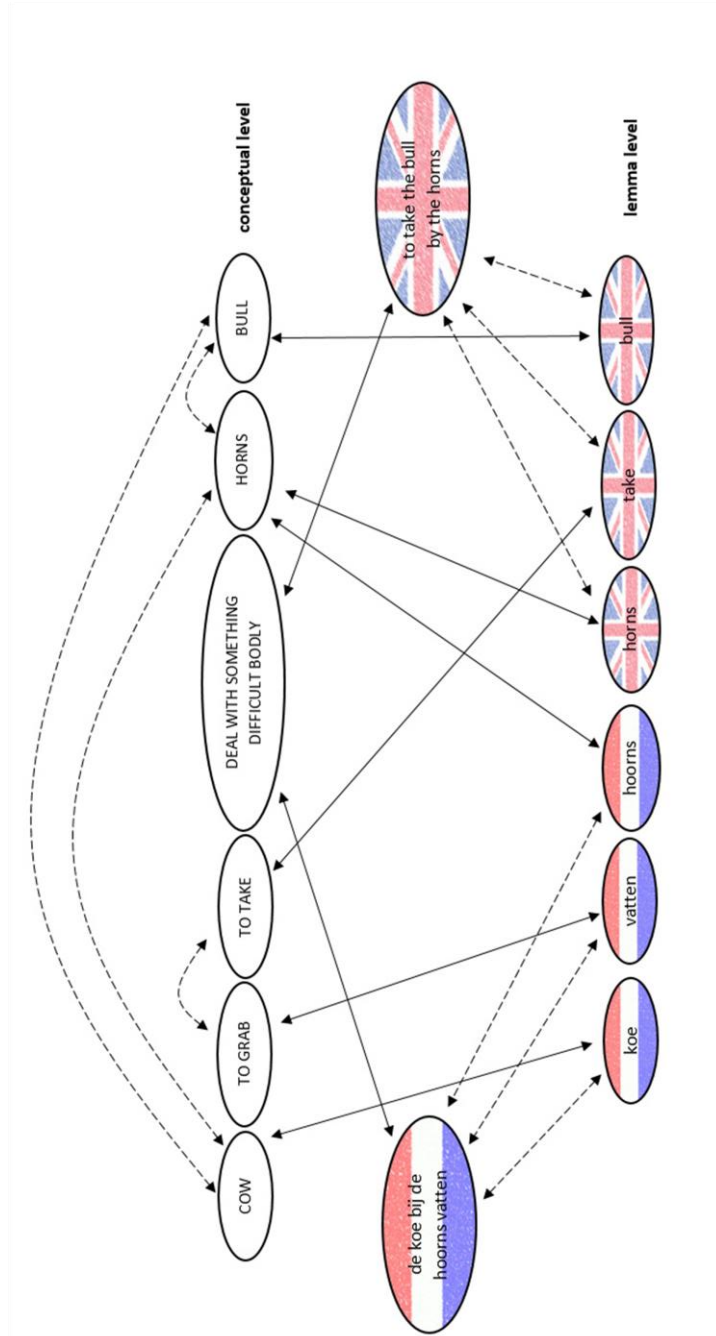


Figure 7.4 Instantiation of the model for an English L2 learner of Dutch in the case of an L2 idiomatic expression that has a meaning equivalent in the L1 in which not all words are direct translations (nW).

The L2 idiom can also have an L1 equivalent in which not all words are direct translations (nW expression). Such a situation is depicted in Figure 7.4. The L2 and L1 superlemma share the same meaning representation and are similar in terms of the event structure and mental model. If L2 learners of Dutch encounter the Dutch idiomatic expression *de koe bij de hoorns vatten* (lit. 'to grab the cow by the horns'), the Dutch simple lemmas may activate the superlemma, which in turn activates the corresponding figurative meaning if L2 learners are familiar with the Dutch idiom. However, if the L2 idiom representation is relatively weak, L2 learners may resort to a different route to access the idiom meaning. The English simple lemma *horns* may be activated by the Dutch simple lemma *hoorns* via the conceptual level. This is similar to the situation presented in Figure 7.3. This, however, is different for the other component words, because they do not share the same concept node. Because, for example, *cow* and *bull* are semantically related, activation of the concept node of *cow* may co-activate the concept of *bull*. This can in turn activate the simple lemma *bull*, leading to activation of the idiomatic meaning via the English superlemma. However, this process is even more indirect than in the case of a translation equivalent as presented in Figure 7.3. As a result, the L2 learners' performance on the idiom knowledge test is not better in the case of nW expressions as compared to L2 idioms that have no L1 equivalent (NE expressions) (**Chapter 3**). However, repeated practice with L2 idioms, which focused on both form and meaning aspects of the idiom, made the L2 learners realize that the nW expressions in the L1 and the L2 share the same meaning, resulting in an advantage of nW expressions as compared to NE expressions (**Chapter 4**).

Figure 7.2, presented earlier, reflects the situation in the case of the Dutch L2 idiomatic expression *het loodje leggen*, which has an equivalent in English that consists of totally different content words, i.e. *to kick the bucket*. In this situation, the figurative meaning can only be retrieved using the direct route (via the Dutch superlemma). The corresponding English superlemma cannot be activated, because the Dutch and English simple lemmas do not share the same concept node through which the English simple lemmas can be activated. In addition, the Dutch and English idiom component words are semantically unrelated. In other words, activation of the concept *piece of lead* does not

lead to co-activation of the concept *bucket*. Therefore, the superlemma *kick the bucket* cannot be activated via the simple lemma *bucket*. If the L2 learner is not familiar with the Dutch idiom, leading to a weaker idiom representation for the Dutch idiom *het loodje leggen*, the idiom meaning cannot be retrieved from the mental lexicon, thus leading to worse performance on the idiom knowledge test for DW idioms as compared to AW idioms (**Chapters 3 and 4**).

It is also possible that the L2 idiomatic expression has no equivalent at all in the L1. This situation is represented in Figure 7.5. If in this situation L2 learners are not familiar with the Dutch idiom *goed uit de verfkomen* (lit. 'to come well out of the paint'), they cannot rely on their L1 to retrieve the idiom's corresponding figurative meaning. The only strategy they might resort to, is to make use of the relative transparency of this idiom. Based on the literal interpretation presented in a specific context they might arrive at the figurative meaning after all.

We note that, although not investigated in this thesis, cognates probably influence the speed and extent to which the L1 superlemma becomes available (e.g., the Dutch L2 idiom *het ijs breken* and its English L1 equivalent *break the ice*). In addition, the situations above are concerned with Dutch and English, two relatively closely related languages. Many idiomatic expressions can therefore be categorized as AW, and nW expressions, in which the meaning is identical and form overlap is present. However, for languages that are more distant, such as Dutch and Arabic, or English and Chinese, the degree of cross-language overlap is relatively limited (Carrol & Conklin, 2014, 2017). For these language pairs, expressions are mainly categorized as DW and NE expressions, which correspond to the instantiations of the bilingual idiom processing model presented in the Figures 7.2 and 7.4. In the bilingual idiom processing model, language distance is incorporated in the degree of cross-language overlap.



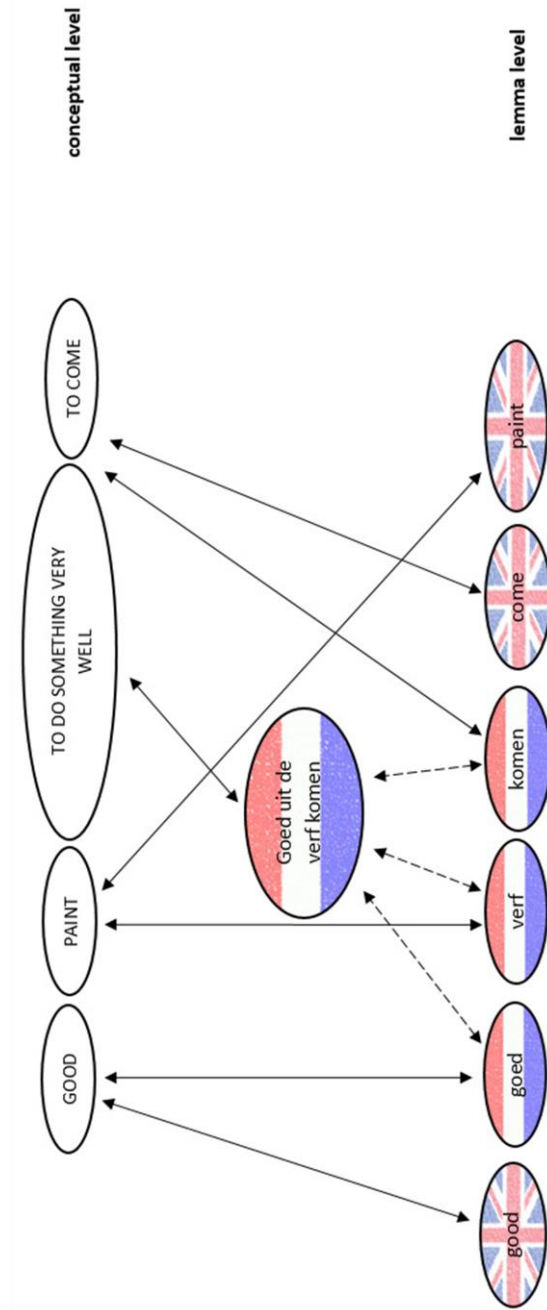


Figure 7.5 Instantiation of the model for an English L2 learner of Dutch in the case of a L2 idiomatic expression that has no equivalent in the L1 (NE).

The different model variants presented above may lead to different RTs and RT patterns during idiom processing. AW idioms, for example, are expected to be processed faster than DW expressions, because AW idioms are identical in form and meaning, while DW idioms are only identical in meaning.

However, no effect of cross-language overlap was found during L2 idiom processing, while this idiom property did affect L2 idiom knowledge (**Chapter 6**). In terms of the bilingual idiom processing model, activation may not have had enough time to flow from the L2 lemmas via the concepts to the L1 lemmas, which in turn activate the L1 superlemma that gives access to the figurative meaning. The idiom knowledge test, however, is an offline task that gives the activation enough time to follow this indirect route.

#### **7.1.4.3 The role of idiom component word characteristics**

In terms of the bilingual idiom processing model, high frequency word forms are expected to activate their corresponding simple lemmas faster than low frequency word forms. As a result, the semantics of the high frequency words also become available faster.

##### *Native speakers*

The bilingual idiom processing model can explain the inhibitory word frequency effect observed during the processing of opaque idioms (**Chapter 5**) in terms of saliency. High frequency words are expected to be more salient compared to the idiom as a whole, while for low frequency words this may be the other way around. If the individual words are of high frequency and thus are more salient than the idiom representation, it may take longer to activate the figurative meaning representation, which is reflected by a slow down during idiom processing. Although the frequency of low frequency words may still be higher than that of idioms, the idiomatic meaning, however, probably is more salient. The simple lemmas may therefore activate the superlemma faster than in the case of frequent words, thus not hindering idiom processing. This relation between frequency and saliency, however, requires further investigation.

*L2 learners*

The facilitatory word frequency effect observed in L2 learners (**Chapter 6**) may be explained in terms of experience with the L2 idiom. In L2 learners, the idiom is relatively weakly represented, which causes less competition between the idiom as a whole and the individual words. In native speakers, on the other hand, the idiom representation is stronger, allowing for more competition between the idiom as a whole and its component words.

By increasing experience with L2 idiomatic expressions, competition may arise between the idiom-final noun and the idiom as a whole. However, the L2 learners did not show such a competition effect after intensively practicing L2 idiomatic expressions through the CALL system (**Chapter 6**). Although this amount of exposure and practice was enough to show positive effects in terms of L2 idiom knowledge, this may not have been enough to lead to a similar sensitivity to individual word frequency as displayed by native speakers.

To summarize, in Section 7.4.1, we presented a bilingual idiom processing model that can account for idiom processing in both native speakers and L2 learners. More experience with idioms leads to stronger idiom representations in both groups and can explain the differences between native speakers and L2 learners observed in this thesis with respect to idiom knowledge and speed of processing. Flow of activation through the model can be different depending on properties of the idiom, such as transparency and, in the case of L2 learners, cross-language overlap. Moreover, the model allows for competition between the idiom component words and the idiom as a whole.

## **7.2 Significance of the research apparatus developed: database, web-based test and CALL system**

As part of this PhD project within the NWO program 'Idiomatic Second Language Acquisition', a number of research instruments were developed that allowed us to conduct research on L2 idiomatic language acquisition and processing that are innovative in various respects.

### 7.2.1 Idiom database

The first of these instruments is a large database of 751 Dutch idiomatic expressions. For 394 expressions, we collected essential information on the properties of these idioms, all obtained from more than 500 native Dutch speakers (see *Chapter 2*). This concerned subjective data on perceived frequency, familiarity, usage, and objective data on corpus frequency and idiom knowledge. An important observation at the beginning of the project was that many studies on idiom processing are based on very small numbers of idioms, for which very limited information is available. To improve on this, we decided to compile this comprehensive database of Dutch idiomatic expressions, for which multiple data on important research variables were subsequently collected and are still being gathered.

This database has been a crucial element in our research. It has provided and will continue to provide useful insights into knowledge, use and intuitions about Dutch idioms by native speakers and language learners, which can be employed to make informed decisions about which idioms to include in research experiments or to make pedagogically sound selections of idioms for learning experiments. We know, for example, which idioms are well known by Dutch native speakers, which ones they find difficult, and which ones they use most often. We also collected information on the extent to which some Dutch idioms exist in other languages like German and Arabic. Since we also collected metadata for all subjects, we also know how knowledge, usage, and subjective ratings of idioms vary as a function of age, educational background, and origin.

The data collection initiative, which was organized as an outreach activity with a test that also allows participants to learn idiomatic expressions (see Section 7.2.2), took place at different editions of a popular language event, the Drongo festival, in 2015 and 2016, the Nijmeegse taalmiddag 2015, and further through the internet. Because this database constitutes a rich resource with potential scientific and societal impact, we decided to make it available to the wider community before the end of the project (Hubers et al., 2018). This database is intended to be a growing collection of Dutch idiomatic expressions and related data to be enriched with subjective ratings and objective data by

other participants beyond those involved in our research, like native speakers of different age groups and education backgrounds and L2 learners with different L1s.

### **7.2.2 Web-based test of idiom knowledge**

A second instrument we designed is a web-based test that was used to collect native speakers' and learners' intuitions about properties of the idioms, as well as objective data on knowledge of idiomatic expressions. This test was used to gather data to include in the idiom database presented in Section 7.2.1 (*Chapters 2 and 3*). Information on cross-language overlap was collected separately. Moreover, the test was used to assess the receptive knowledge of Dutch idiomatic expressions by participants in pre-tests and post-test (*Chapter 4*). This is a considerable improvement compared to previous approaches in which idiom knowledge was either not tested at all, or was estimated based on self-reported data and/or familiarity judgments, or data collected from other, comparable subjects.

In the meantime, this test has been employed as such or in adjusted forms to collect data on knowledge of Dutch idiomatic expressions in other target groups such as L2 learners of Dutch with Arabic as L1, and Dutch emigrants in the diaspora. The Arabic L2 learners of Dutch completed the web-based idiom knowledge test as part of the same CALL-based idiom learning study as reported in *Chapter 4* with German L2 learners of Dutch. For the Dutch emigrants, the test had been adjusted in such a way that only data on idiom knowledge and transparency intuitions were collected. These data have been collected as part of another research project. The data obtained from Arabic learners and Dutch emigrants have not been analyzed in the framework of the research reported on in this thesis, but are readily available for future studies.

### **7.2.3 CALL system**

A third instrument we designed is a Computer Assisted Language Learning system that provides learners with the opportunity to practice Dutch idiomatic expressions and receive automatic feedback from the computer. Employing a CALL system for research on idiom learning has

several important advantages. First, a number of crucial variables such as language materials, type and intensity of practice and type, amount and timing of feedback can be controlled, systematically varied and uniformly provided in a way that would not be possible in classroom situations. However, this does not mean the experimental environment thus created becomes so unrealistic that it could not be reproduced in normal teaching situations. On the contrary, a second important advantage is that a CALL system environment has high ecological validity as CALL-based practice could be easily incorporated in regular language education as a complement to teacher fronted instruction. A third benefit of using a CALL system is that learners can work independently at their own level and pace and can receive a degree level of intensive and individualized practice that would not otherwise be feasible in classroom instruction.

For this CALL system, we designed training sessions intended to practice Dutch idiomatic expressions through a variety of paradigms and exercises (see *Chapter 4*). In a first training session, L2 learners participated in a Paired Associated Learning (PAL) paradigm (Steinel et al., 2007), in which Dutch idiomatic expressions were briefly presented, one by one, together with their figurative meaning. Each idiom was presented at the center of the screen with the corresponding meaning directly below for 30 seconds. Subsequently, the next idiom-meaning pair would appear automatically on the screen. Participants were instructed to carefully read the idioms and their meanings, but did not have to perform any other task.

A second type of training comprised gap-fill exercises in which crucial words in the idioms had to be filled in. A third exercise type consisted of sentence completion tasks in which given sentences had to be finished off with the appropriate idiomatic expression. A fourth form of practice comprised idiom selection exercises in which a definition of the figurative meaning had to be matched to the corresponding idiom.

In the gap-fill, sentence completion, and idiom selection exercises, automatic corrective feedback was provided immediately after the participants had typed in the answer. The feedback showed both the correct answer and the answer as provided by the learner. In addition, in all these exercises it was possible for learners to repeat the idiomatic expressions silently or in a read-aloud mode. More details about the

exercises can be found in the screenshots in the Supplementary Materials (Figures S4 – S10).

An important aspect of this CALL system is that it logs all system-learner interactions, to allow for more insight into learner behavior. All utterances by the users are recorded and are thus available for subsequent research. The interactions are stored in a database and this allows us to look in detail at learner behavior and inspect the logs for irregular behavior. We store interaction data that can be relevant for research purposes, for instance for studying the effects of practice and corrective feedback on performance and proficiency. The logged interaction data were not a topic of study in the research reported on in this thesis, but are available for future research.

All the instruments presented above, database, web-based test, CALL system and exercises are available not only to conduct further research on Dutch idiom learning, but also to develop pedagogically sound didactic materials.

### **7.3 Relevance for scientific research**

The research reported on in this thesis provides important insights into the multifaceted nature of idiomatic expressions. Since idiomatic expressions have been investigated in both the L1 and the L2 from the perspective of knowledge, learning, representation and processing (**Chapters 2, 3, 4, 5 and 6**), this research can be seen as the first attempt to connect these different perspectives to obtain a comprehensive model of L1 and L2 idiom processing, with an important role for L2 idiom experience which has been experimentally tested.

A scientifically important feature of this thesis is the adoption of an integrated approach linking idiom acquisition in the L2 to on-line idiom processing in the same participants using techniques from both applied linguistics and psycholinguistics (**Chapter 6**). The integration into one coherent approach helps broaden and deepen our understanding of idiom processing in the L2, thus answering the call by Wray (2009, p. 2) “to explore the fullest range of opportunities for making a useful contribution to work in this field”.

This research also enables an analysis of linguistic item learning (**Chapter 4**) and on-line speech processing (**Chapters 5 and 6**) beyond the word level. Studying L2 idiom learning with respect to on-line production in the same participants in naturalistic teaching/training situations provides challenging data for computational human speech processing, which so far has mainly focused on controlled word processing in restricted laboratory conditions.

#### **7.4 Relevance for teaching practice**

The research conducted in this thesis not only has scientific value, it also provides important insights for teaching practice. Idiomatic language is an aspect of an L2 that usually does not receive much attention in L2 classrooms. However, L2 learners do consider learning L2 idioms important (Liontas, 2015a), and L2 learners come across as more proficient if they use idiomatic language (Boers, Eyckmans, Kappel, et al., 2006).

**Chapter 4** shows that through the use of a properly designed CALL system that provides immediate feedback, L2 learners are able to learn L2 idiomatic expressions up to a level that is comparable to native performance in terms of meaning recognition. More specifically, it shows that L2 learners can learn even from only two idiom presentations and with intensive practice they are able to correctly recognize the meaning of idiomatic expressions to the same extent as native speakers. Therefore, we encourage teachers to pay attention to idiomatic language in the L2 classroom and we can provide specific information on how intensively idioms should be practiced based on their distinguishing properties.

This thesis provides information on the properties that make L2 idioms more difficult or easier to learn and how this knowledge could be exploited in idiom teaching. **Chapters 3 and 4** show that transparency and cross-language overlap are important predictors for L2 idiom knowledge. In **Chapter 4**, however, this information was especially used after L2 learners completed the training. This suggests that L2 learners are not necessarily aware of the usefulness of these properties from the start. We therefore recommend teachers to direct the L2 learner's attention to transparency and cross-language overlap information when introducing L2 idiomatic expressions. Moreover, specific attention



should be paid to opaque idiomatic expressions, to idioms that do not exist in the learner's L1, and to idioms that do exist in the L1, but with a completely different form. Last, but not least, this research has shown that employing a CALL system is a commendable way of organizing and facilitating L2 idiom practice.

## **7.5 Methodological recommendations**

This thesis addressed a number of methodological issues that are important for future research on idiomatic expressions. The issues concern norming (Section 7.5.1) and research design in L2 idiom processing research (Section 7.5.2).

### **7.5.1 Norming studies**

As idiom properties are found to influence idiom knowledge, learning and processing, many researchers investigating idiomatic expressions obtain information on these idiom properties through norming studies. In these studies, native speakers are usually asked to rate idiomatic expressions on a number of dimensions. The question then is to what extent participants are able to reliably judge these aspects of idiomatic expressions. Only a few studies investigated the reliability of the ratings and obtained mixed results. In *Chapter 2*, we reported on a large-scale rating study with native speakers in which we explicitly investigated the reliability of their subjective judgements using different measures. By carefully operationalizing the idiom properties, the intuitions of native speakers about these properties turned out to be highly reliable. However, it is important to use the appropriate metric to measure reliability. Previous studies that did address reliability used different metrics, such as Cronbach's alpha, Intraclass Correlation Coefficient (ICC), and Krippendorff's alpha. We compared these measures and the D-coefficient, a measure that comes from Generalizability Theory.

Our reliability analysis led us to recommend that future norming studies use the D-coefficient as a measure of reliability, because the D-coefficient can handle all kinds of research designs and measurement levels, and it allows for generalization across raters. If the research design allows, the ICC is also an appropriate reliability metric, leading to results

comparable to that of the D-coefficient. The advantage of the D-coefficient over the ICC is that it allows for assessment of the minimum number of raters that are required to obtain reliable data. *Chapter 2* suggests that for experience-based dimensions that are carefully operationalized, 10 raters might be sufficient to obtain reliable data, whereas for judgements of content-based dimensions to be reliable at least 20 participants are required.

Subjective judgements about idiom properties are often also used in research on idiomatic expressions in L2 learners. This information is, for example, used to select materials for experiments, or to assess the effect of idiom properties on L2 idiom learning and processing. Sometimes intuitions about these idiom properties are obtained from L2 learners, but more often the intuitions by native speakers are used in L2 idiom research. If these intuitions are obtained from L2 learners the question arises to what extent L2 learners are in fact capable of developing reliable intuitions about idiom properties. If the intuitions are obtained from native speakers, one might wonder whether this properly reflects the L2 learner's ideas about L2 idioms. In *Chapter 3*, we investigated these questions in a rating study similar to the one reported in Chapter 2. L2 learners were found to be capable of developing reliable intuitions about idiom properties. The precise and careful way in which the questions about the idiom properties were formulated probably contributed to the high reliability. In addition, the L2 intuitions about familiarity, transparency and imageability were more informative when studying L2 idiom knowledge than the corresponding L1 intuitions.

Therefore, we recommend that researchers in L2 idiom processing and idiom learning collect and use ratings of idiomatic expressions obtained from L2 learners rather than from native speakers. When collecting these ratings, however, it is important to precisely formulate the rating questions in order to obtain reliable results.

### **7.5.2 Research on L2 idiom processing**

Relatively recently, researchers started to address idiom processing in L2 learners, and the extent to which different processes underlie idiom processing in native speakers and L2 learners (Beck & Weber, 2016a; Carrol & Conklin, 2014, 2017, Cieślicka, 2006, 2013; van Ginkel &

Dijkstra, 2019). To gain insight into L2 idiom representation, experimental techniques have been applied that directly tap into the processing of idiomatic expressions during L2 comprehension. An example are reaction time data, which have often been employed, because they provide more objective measures of familiarity and can thus reveal important processing details. By means of the cross-modal priming technique, for example, reaction time data can be obtained that can inform researchers about the role of individual literal word meanings during idiom processing.

However, since L2 learners experience enormous difficulties with idiomatic expressions in the L2 (Cieślicka, 2006; Conklin & Schmitt, 2008; Ellis et al., 2008; Kovecses & Szabó, 1996; Wray, 2000) and idioms often go unnoticed by L2 learners (Boers & Lindstromberg, 2012; Eyckmans et al., 2007; Jones & Haywood, 2004), researchers should know to what extent the L2 participants in idiom processing experiments are familiar with the idiomatic expressions under study. This way, differences between L1 and L2 idiom processing cannot be ascribed to the fact that L2 learners interpret an idiom as any other literal phrase. To verify whether L2 learners know the idioms, researchers often ask the L2 learners to rate the idioms on familiarity after the psycholinguistic experiment. This, however, only gives very general information about the L2 idiom knowledge of the participants and many expressions may be excluded from the analysis afterwards.

One way to resolve this issue is to have L2 learners undergo a training beforehand. This way, researchers can ensure that the L2 participants know the idiom before they take part in psycholinguistic experiments. In *Chapter 6* we adopted such a design. By combining techniques from both applied linguistics and psycholinguistics, the participants all had the same minimum knowledge of the idiomatic expressions to be tested. Effects in L2 idiom processing could therefore not be attributed to lack of prior knowledge, but were more likely to be due to the effect of the factors under investigation. Moreover, this design allowed us to address new questions with respect to the effect of exposure or practice on L2 idiom processing.

We therefore recommend researchers in the area of L2 idiom processing to carefully assess the L2 participants' idiom knowledge prior

to the experiment, by, for example, using an extensive familiarization phase or focused training. However, the extent to which all idioms under study should be included in such a training may depend on the specific research question of the researcher.

## 7.6 Perspectives for future research

In *Chapters 2 and 3*, we collected information on idiom knowledge by asking participants to identify the correct meaning of the idiomatic expressions out of four alternatives (idiom meaning recognition), and by asking participants to provide the meaning in an open question (idiom meaning recall). So far, however, we only analyzed the data on the idiom meaning recognition in relation to the idiom properties. Although this is already a big step forward since the majority of rating studies previously conducted subjectively assessed idiom knowledge, the data on the idiom meaning recall may provide interesting additional information. It would be worth investigating to what extent idiom meaning recall is different from idiom meaning recognition, and whether idiom properties differently affect recognition and recall.

In addition, imageability was found to negatively affect receptive idiom knowledge in both native speakers and L2 learners (*Chapters 2 and 3*). This negative effect may have been due to a lack of specificity in operationalization, because it is hard to determine whether participants formed an image of the literal or figurative interpretation. Future research is needed to clarify this issue.

In *Chapter 2*, we mainly collected data from native speakers between 18 and 30 years old. However, older people also completed our web-based idiom knowledge test. Although the number of older people that filled in the test was too limited to include them in the final analysis, preliminary analyses suggest that age is an important determinant of idiom knowledge. An interesting avenue for further research is, therefore, to investigate the effect of age on idiom knowledge and processing.

In *Chapter 4*, we investigated idiom learning by German L2 learners of Dutch. More specifically, we manipulated intensity of practice and modality of practice. As expected, intensive practice facilitated L2 idiom learning. However, we did not find an effect of reading modality.

We hypothesized that repeating the sentences after completing the exercises by reading aloud would enhance idiom learning as compared to reading these sentences silently. The implementation of our reading modality manipulation may not have been optimal to show any effect. More research is required to shed light on the lack of a positive effect of reading aloud. It may be worthwhile for future research to include idiom production exercises in order to examine this issue. In fact, in the current version of our CALL system we incorporated automatic speech recognition. This way, L2 learners can orally complete the exercises and immediate feedback can still be provided.

With respect to idiom processing, we investigated native speakers (*Chapter 5*) and L2 learners (*Chapter 6*) using a paradigm inspired by Rommers, Dijkstra, & Bastiaansen (2013). While the results obtained from native speakers in Experiment 2 could be interpreted, this was different for the results from L2 learners. Presenting the target word with a delay of 200 ms worked in the case of the native speakers, but did not lead to the expected pattern of results in our baseline condition (the literal context) with L2 learners. Future research is necessary to clarify our results on this point.

In *Chapter 6*, we also investigated the effect of practice on L2 idiom processing. Although the effect of practice was not significant, a trend was visible in the expected direction. L2 learners were faster in response to the idiom-final nouns of intensively practiced idioms than to idiom-final nouns of idioms that had received limited practice only. Increasing the amount of practice, and providing even more engaging and demanding practice may result in differences in L2 idiom processing in the future.

This thesis investigated German L2 learners of Dutch only. As we mentioned in the introduction chapter, this was a logical starting point for our research, but it is clear that other options could be explored. For instance, it would be interesting to conduct similar experiments with an L1 and L2 that lie further apart. Preliminary learning experiments with Arabic L2 learners of Dutch showed that for them transparency was an important determinant of idiom learning, which is consistent with the findings for German L2 learners of Dutch. However, since the degree of cross-language overlap between Arabic and Dutch is relatively low, this

factor did not seem to facilitate idiom learning, while the German L2 learners were able to use this property. In addition, in this thesis, we assessed receptive idiom knowledge one week after the L2 learners completed the learning phase. It would be interesting to investigate how durable the learning is by including a delayed post-test. It would be worthwhile to further investigate these issues.

To summarize, this thesis provides several useful insights into L2 idiom knowledge, learning, and processing, but it is clear that a number of important issues are still unresolved. *Chapter 4* suggests that after sufficient exposure the idiom knowledge of native speakers and that of L2 learners is comparable. However, the psycholinguistic experiments reveal differences between native speakers and L2 learners (*Chapters 5 and 6*). Although these results too indicate an important role for exposure, further research is needed to shed more light on this issue. Experiments that systematically vary exposure and practice, but also the time course in L2 idiom processing seem to be the way forward.

## 7.7 Conclusions

This thesis investigated the knowledge, representation, and processing of idiomatic expressions by L2 learners and native speakers. We have shown that differences in receptive idiom knowledge between native speakers and L2 learners are mainly due to differences in experience with idiomatic expressions, and that idiom properties such as transparency and imageability similarly affect this type of idiom knowledge in native speakers and L2 learners. Moreover, this thesis has found that differences in idiom knowledge can be overcome by providing focused training through a CALL-system in which L2 learners received immediate corrective feedback.

In terms of idiom processing, this thesis has revealed that the amount of exposure to L2 idioms during the CALL-based training was not sufficient to resolve differences between native speakers and L2 learners, but that even more exposure can lead to similar results.

We have integrated the insights obtained in this thesis into a newly proposed bilingual idiom processing model that is able to account for both L1 and L2 idiom processing, with a special role for experience, idiom

properties, and individual word properties. In all, this thesis has shown that native speakers and L2 learners may be chalk and cheese with respect to their experience with idiomatic expressions, but that in terms of their underlying architecture they are two of a kind.





## References

- Abel, B. (2003). English idioms in the first language and second language lexicon: a dual representation approach. *Second Language Research*, 19(4), 329–358. <http://doi.org/10.1191/0267658303sr226oa>
- Ahmadi, A., Sahragard, R., & Babaie Shalmani, H. (2017). Anthropomorphism – matters or not? On agent modality and its implications for teaching English idioms and design decisions. *Computer Assisted Language Learning*, 30(1–2), 149–172. <http://doi.org/10.1080/09588221.2017.1284132>
- Aijmer, K. (1996). *Conversational Routines in English*. London and New York: Longman.
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <http://doi.org/10.18637/jss.v067.i01>
- Beck, S. D., & Weber, A. (2016a). Bilingual and monolingual idiom processing is cut from the same cloth: The role of the L1 in literal and figurative meaning activation. *Frontiers in Psychology*, 7, 1–16. <http://doi.org/10.3389/fpsyg.2016.01350>
- Beck, S. D., & Weber, A. (2016b). English-German Database of Idiom Norms (DIN).
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Bobrow, S. a, & Bell, S. M. (1973). On catching on to idiomatic expressions. *Memory & Cognition*, 1(3), 343–346. <http://doi.org/10.3758/BF03198118>
- Boers, F., Demecheleer, M., & Eyckmans, J. (2004). Cross-cultural Variation as a Variable in Comprehending and Remembering Figurative Idioms. *European Journal of English Studies*, 8(3), 375–388. <http://doi.org/10.1080/1382557042000277449>
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: putting a Lexical Approach to the test. *Language Teaching Research*, 10(3), 245–261. <http://doi.org/10.1191/1362168806lr195oa>
- Boers, F., Eyckmans, J., & Stengers, H. (2006). Motivating multiword units: Rationale, mnemonic benefits, and cognitive style variables. In S. H. Foster-Cohen, M. Medved Krajnovic, & J. Mihaljević Djigunović (Eds.), *EUROSLA Yearbook* (Vol. 6, pp. 169–190). Amsterdam/Philadelphia: John Benjamins Publishing Company. <http://doi.org/10.1075/eurosla.6.11boe>

- Boers, F., Eyckmans, J., & Stengers, H. (2007). Presenting figurative idioms with a touch of etymology: more than mere mnemonics? *Language Teaching Research*, 11(1), 43–62. <http://doi.org/10.1177/1362168806072460>
- Boers, F., & Lindstromberg, S. (2012). Experimental and Intervention Studies on Formulaic Sequences in a Second Language. *Annual Review of Applied Linguistics*, 32, 83–110. <http://doi.org/10.1017/S0267190512000050>
- Boers, F., Lindstromberg, S., Littlemore, J., Stengers, H., & Eyckmans, J. (2008). Variables in the mnemonic effectiveness of pictorial elucidation. In F. Boers & S. Lindstromberg (Eds.), *Cognitive Linguistic approaches to teaching vocabulary and phraseology* (pp. 189–216). Berlin, Germany: Mouton De Gruyter.
- Boers, F., Píríz, A. M. P., Stengers, H., & Eyckmans, J. (2009). Does pictorial elucidation foster recollection of idioms? *Language Teaching Research*, 13(4), 367–382. <http://doi.org/10.1177/1362168809341505>
- Boers, F., & Webb, S. (2015). Gauging the Semantic Transparency of Idioms: Do Natives and Learners See Eye to Eye? In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Figurative Language Processing* (pp. 368–392). New York, NY: Cambridge University Press.
- Bonin, P., Méot, A., & Bugajska, A. (2013). Norms and comprehension times for 305 French idiomatic expressions. *Behavior Research Methods*, 45, 1259–1271. <http://doi.org/10.3758/s13428-013-0331-4>
- Boon, C. A. den, & Hendrickx, R. (2017). Dikke Van Dale Online 2015 t/m 2017. Retrieved from <http://www.vandale.nl/>
- Brennan, R. L. (2001). *Generalizability theory*. New York: Springer.
- Brysbaert, M., Mander, P., & Keuleers, E. (2018). The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science*, 27(1), 45–50. <http://doi.org/10.1177/0963721417727521>
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3), 904–911. <http://doi.org/10.3758/s13428-013-0403-5>
- Cacciari, C. (2014). Processing multiword idiomatic strings: Many words in one? *The Mental Lexicon*, 9(2), 267–293. <http://doi.org/10.1075/ml.9.2.05cac>
- Cacciari, C., & Glucksberg, S. (1991). Understanding idiomatic expressions: The contribution of word meanings. In G. B. Simpson

- (Ed.), *Understanding Word and Sentence* (pp. 217–240). Amsterdam: North-Holland. [http://doi.org/10.1016/S0166-4115\(08\)61535-6](http://doi.org/10.1016/S0166-4115(08)61535-6)
- Cacciari, C., & Glucksberg, S. (1995). Understanding idioms: Do visual images reflect figurative meanings? *European Journal of Cognitive Psychology*, 7(3), 283–305. <http://doi.org/10.1080/09541449508402450>
- Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, 27, 668–683. <http://doi.org/10.1007/BF01067624>
- Caillies, S. (2009). Descriptions de 300 expressions idiomatiques : familiarité, connaissance de leur signification, plausibilité littérale, «décomposabilité» et «prédictibilité». *L'Année Psychologique*, 109(3), 463–508. <http://doi.org/10.4074/S0003503309003054>
- Carrol, G., & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition*, 17(4), 784–797. <http://doi.org/10.1017/S1366728913000795>
- Carrol, G., & Conklin, K. (2017). Cross language lexical priming extends to formulaic units: Evidence from eye-tracking suggests that this idea “has legs.” *Bilingualism: Language and Cognition*, 20(2), 299–317. <http://doi.org/10.1017/S1366728915000103>
- Carrol, G., Littlemore, J., & Gillon Dowens, M. (2017). Of false friends and familiar foes: Comparing native and non-native understanding of figurative phrases. *Lingua*. <http://doi.org/10.1016/j.lingua.2017.11.001>
- Cermak, L., & Craik, F. I. M. (1979). *Levels of processing in human memory*. (L. Cermak & F. I. M. Craik, Eds.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Charteris-Black, J. (2002). Second Language Figurative Proficiency: A Comparative Study of Malay and English. *Applied Linguistics*, 23, 104–133.
- Cieślicka, A. B. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research*, 22(2), 115–144. <http://doi.org/10.1191/0267658306sr263oa>
- Cieślicka, A. B. (2013). Do Nonnative Language Speakers Chew the Fat and Spill the Beans with Different Brain Hemispheres? Investigating Idiom Decomposability with the Divided Visual Field Paradigm. *Journal of Psycholinguistic Research*, 42(6), 475–503. <http://doi.org/10.1007/s10936-012-9232-4>
- Cieślicka, A. B. (2015). Idiom Acquisition and Processing by

- Second/Foreign Language Learners. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Figurative Language Processing* (pp. 208–244). New York, NY: Cambridge University Press.
- Citron, F. M. M., Cacciari, C., Kucharski, M., Beck, L., Conrad, M., & Jacobs, A. M. (2016). When emotions are expressed figuratively: Psycholinguistic and Affective Norms of 619 Idioms for German (PANIG). *Behavior Research Methods*, *48*(1), 91–111. <http://doi.org/10.3758/s13428-015-0581-4>
- Conklin, K., & Schmitt, N. (2008). Formulaic Sequences: Are They Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers? *Applied Linguistics*, *29*(1), 72–89. <http://doi.org/10.1093/applin/amm022>
- Cooper, T. C. (1999). Processing of Idioms by L2 Learners of English. *TESOL Quarterly*, *33*(2), 233–262. <http://doi.org/10.2307/3587719>
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. [http://doi.org/10.1016/S0022-5371\(72\)80001-X](http://doi.org/10.1016/S0022-5371(72)80001-X)
- Cronk, B. C., Lima, S. D., & Schweigert, W. A. (1993). Idioms in sentences: Effects of frequency, literalness, and familiarity. *Journal of Psycholinguistic Research*, *22*(1), 59–82. <http://doi.org/10.1007/BF01068157>
- Cutting, J. C., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory & Cognition*, *25*(1), 57–71.
- de Bot, K. (1992). A Bilingual Production Model: Levelt's "Speaking" Model Adapted. *Applied Linguistics*, *13*(1), 1–24. <http://doi.org/10.1093/applin/13.1.1>
- de Bot, K., Cox, A., Ralston, S., Schaufeli, A., & Weltens, B. (1995). Lexical processing in bilinguals. *Second Language Research*, *11*(1), 1–19. <http://doi.org/10.2307/43103440>
- De Deyne, S., & Storms, G. (2008). Word associations : Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods*, *40*(1), 198–205. <http://doi.org/10.3758/BRM.40.1.198>
- de Vet, H. C. W., Terwee, C. B., Knol, D. L., & Bouter, L. M. (2006). When to use agreement versus reliability measures. *Journal of Clinical Epidemiology*, *59*(10), 1033–1039. <http://doi.org/10.1016/J.JCLINEPI.2005.10.015>
- Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*(2), 163–188.
- Duyck, W., Anseel, F., Szmalec, A., Mestdagh, P., Tavernier, A., &

- Hartsuiker, R. J. (2008). Improving accuracy in detecting acoustic onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 34(5), 1317–26. <http://doi.org/10.1037/0096-1523.34.5.1317>
- Ellis, N. C. (2001). Memory for Language. In P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press.
- Ellis, N. C. (2012). Formulaic Language and Second Language Acquisition: Zipf and the Phrasal Teddy Bear. *Annual Review of Applied Linguistics*, 32, 17–44. <http://doi.org/10.1017/S0267190512000025>
- Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL. *TESOL Quarterly*, 42(3), 375–396.
- Ellis, N. C., & Wulff, S. (2015). Second Language Acquisition. In E. Dabrowska & D. Divjak (Eds.), *Handbook of Cognitive Linguistics* (pp. 409–716). Mouton De Gruyter.
- Elshoff, J. (2014). Designing a Task Based Curriculum for Intensive Language Training. In S. Jager, L. Bradley, E. J. Meima, & S. Thouësny (Eds.), *CALL Design: Principles and Practice; Proceedings of the 2014 EUROCALL Conference* (pp. 66–71). Dublin: Research-publishing.net. <http://doi.org/10.14705/rpnet.2014.000196>
- Erman, B. (2007). Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics*, 12(1), 25–53. <http://doi.org/10.1075/ijcl.12.1.04erm>
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20(1), 29–62. <http://doi.org/10.1515/text.1.2000.20.1.29>
- Eyckmans, J., Boers, F., & Stengers, H. (2007). Identifying chunks: Who can see the wood for the trees? *Language Forum*, 33(2), 85–100. Retrieved from [https://cris.vub.be/en/publications/identifying-chunks-who-can-see-the-wood-for-the-trees\(093b790a-e2bc-4b9f-b2b1-f4bdac81096b\).html](https://cris.vub.be/en/publications/identifying-chunks-who-can-see-the-wood-for-the-trees(093b790a-e2bc-4b9f-b2b1-f4bdac81096b).html)
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491–505. <http://doi.org/10.1111/j.1469-8986.2007.00531.x>
- Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, 41(4), 469–495. <http://doi.org/10.1006/jmla.1999.2660>
- Federmeier, K. D., McLennan, D. B., Ochoa, E., & Kutas, M. (2002). The

- impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, 39(2), 133–146. <http://doi.org/10.1111/1469-8986.3920133>
- Fox, J. (2003). Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software*, 8(15), 1–27. <http://doi.org/10.2307/271037>
- García, O., Cieślicka, A. B., & Heredia, R. R. (2015). Nonliteral Language Processing and Methodological Considerations. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Figurative Language Processing* (pp. 117–168). New York: Cambridge University Press. <http://doi.org/10.1017/CBO9781139342100.009>
- Geeraerts, D. (1995). Specialization and reinterpretation in idioms. In M. Everaert, E. J. van der Linden, A. Schenk, & R. Schreuder (Eds.), *Idioms: Structural and psychological perspectives* (pp. 57–73). Mahwah, NJ: Erlbaum.
- Genootschap OnzeTaal. (2017). Uitdrukkingen en Spreekwoorden. Retrieved from <https://onzetaal.nl/taaladvies/thematisch/uitdrukkingen-en-spreekwoorden/>
- Gernsbacher, M. A. (1984). Resolving 20 Years of Inconsistent Interactions Between Lexical Familiarity and Orthography, Concreteness, and Polysemy. *Journal of Experimental Psychology. General*, 113(2), 256–281. <http://doi.org/10.1016/j.jacc.2007.01.076>. White
- Gibbs, R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition*, 8(2), 149–156.
- Gibbs, R. W., Bogdanovich, J. M., Sykes, J. R., & Barr, D. J. (1997). Metaphor in Idiom Comprehension. *Journal of Memory and Language*, 37(2), 141–154. <http://doi.org/10.1006/JMLA.1996.2506>
- Gibbs, R. W., & Nayak, N. P. (1989). Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive Psychology*, 21(1), 100–138. [http://doi.org/10.1016/0010-0285\(89\)90004-2](http://doi.org/10.1016/0010-0285(89)90004-2)
- Gibbs, R. W., Nayak, N. P., Bolton, J. L., & Keppel, M. E. (1989). Speakers' assumptions about the lexical flexibility of idioms. *Memory & Cognition*, 17(1), 58–68.
- Gibbs, R. W., Nayak, N. P., & Cutting, C. (1989). How to Kick the Bucket and Not Decompose: Analyzability and Idiom Processing. *Journal of Memory and Language*, 28, 576–593.
- Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, 8(3), 183–206. <http://doi.org/10.1515/cogl.1997.8.3.183>
- Giora, R. (2002). Literal vs. figurative language: Different or equal?

- Journal of Pragmatics*. [http://doi.org/10.1016/S0378-2166\(01\)00045-5](http://doi.org/10.1016/S0378-2166(01)00045-5)
- Glucksberg, S. (1993). Idiom meaning and allusional content. In C. Cacciari & P. Tabossi (Eds.), *Idioms Processing, Structure and Interpretation* (pp. 3–26). Hillsdale: Erlbaum.
- Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: the frequency-lag hypothesis. *Journal of Experimental Psychology. General*, *140*(2), 186–209. <http://doi.org/10.1037/a0022256>
- Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311–343. <http://doi.org/10.1177/0267658312461497>
- Gries, S. T., & Ellis, N. C. (2015). Statistical Measures for Usage-Based Linguistics. *Language Learning*, *65*(S1), 228–255. <http://doi.org/10.1111/lang.12119>
- Holcomb, P. J., Reder, L., Misra, M., & Grainger, J. (2005). The effects of prime visibility on ERP measures of masked priming. *Brain Research. Cognitive Brain Research*, *24*(1), 155–72. <http://doi.org/10.1016/j.cogbrainres.2005.01.003>
- Hsu, J. T., & Chiu, C. (2008). Lexical Collocations and their Relation to Speaking Proficiency of College EFL Learners in Taiwan. *The Asian EFL Journal*, *10*(1), 181–204.
- Hubers, F., Cucchiari, C., Strik, H., & Dijkstra, T. (2019). Normative Data of Dutch Idiomatic Expressions: Subjective Judgments You Can Bank on. *Frontiers in Psychology*, *10*, 1–15. <http://doi.org/10.3389/fpsyg.2019.01075>
- Hubers, F., van Ginkel, W., Cucchiari, C., Dijkstra, T., & Strik, H. (2016). How Native Speakers see the Light. In *Paper presented at FLaRN 2016*.
- Hubers, F., van Ginkel, W., Cucchiari, C., Strik, H., & Dijkstra, T. (2018). Normative data on Dutch idiomatic expressions: Native speakers. DANS [Dataset]. <http://doi.org/https://doi.org/10.17026/dans-zjx-hnsk>
- Irujo, S. (1986a). A piece of cake: learning and teaching idioms. *ELT Journal*, *40*(3), 236–243. <http://doi.org/10.1093/elt/40.3.236>
- Irujo, S. (1986b). Don't Put Your Leg in Your Mouth: Transfer in the Acquisition of Idioms in a Second Language. *TESOL Quarterly*, *20*(2), 287–304.
- Irujo, S. (1993). Stearing Clear: Avoidance in the Production of Idioms.

- International Review of Applied Linguistics in Language Teaching*, 31(3), 205–219.
- Isobe, Y. (2011). Representation and processing of formulaic sequences in L2 mental lexicon: how do Japanese EFL learners process multi-word expressions? *JACET Kansai Journal*, (13), 38–49. Retrieved from <https://ci.nii.ac.jp/naid/110008608532>
- Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *Modern Language Journal*, 91(3), 433–445. <http://doi.org/10.1111/j.1540-4781.2007.00589.x>
- Jones, M. A., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, processing and use* (pp. 269–300). John Benjamins Publishing Company. <http://doi.org/10.1075/llt.9.14jon>
- Kecskes, I. (2006). On my mind: thoughts about salience, context and figurative language from a second language perspective. *Second Language Research*, 22(2), 1–19.
- Kecskes, I. (2007). Formulaic language in English Lingua Franca. In I. Kecskes & L. R. Horn (Eds.), *Explorations in pragmatics: Linguistic, cognitive and intercultural aspects* (pp. 191–219). Berlin/New York: Mouton de Gruyter.
- Kecskes, I. (2015). Is the Idiom Principle Blocked in Bilingual L2 Production? In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Figurative Language Processing* (pp. 28–52). New York, NY: Cambridge University Press.
- Kecskés, I. (2000). A cognitive-pragmatic approach to situation-bound utterances. *Journal of Pragmatics*, 32(5), 605–625. [http://doi.org/10.1016/S0378-2166\(99\)00063-6](http://doi.org/10.1016/S0378-2166(99)00063-6)
- Kellerman, E. (1979). Transfer and Non-Transfer: Where We Are Now. *Studies in Second Language Acquisition*, 2(1), 37–57. Retrieved from <http://journals.cambridge.org/action/displayFulltext?type=1&fid=2543436&jid=SLA&volumeId=2&issueId=01&aid=2543428>
- Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, 42(3), 643–650. <http://doi.org/10.3758/BRM.42.3.643>
- Keysar, B., & Bly, B. (1995). Intuitions of the Transparency of Idioms: Can One Keep a Secret by Spilling the Beans? *Journal of Memory and Language*, 34(1), 89–109. <http://doi.org/10.1006/JMLA.1995.1005>
- Kim, C. (2016). L2 learners' recognition of unfamiliar idioms composed of familiar words. *Language Awareness*, 25(1–2), 89–109.



- <http://doi.org/10.1080/09658416.2015.1122025>
- Kovecses, Z., & Szabó, P. (1996). Idioms: A View from Cognitive Semantics. *Applied Linguistics*, 17(3), 326–255.
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13). <http://doi.org/10.18637/jss.v082.i13>
- Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition* (pp. 20–34). Cambridge: Cambridge University Press. <http://doi.org/10.1017/CBO9781139524643.004>
- Laufer, B. (2000). Avoidance of idioms in a second language: The effect of L1-L2 degree of similarity. *Studia Linguistica*, 54(2), 186–196. <http://doi.org/10.1111/1467-9582.00059>
- Lee, H.-W., Rayner, K., & Pollatsek, A. (1999). The time course of phonological, semantic, and orthographic coding in reading: Evidence from the fast-priming technique. *Psychonomic Bulletin & Review*, 6(4), 624–634. <http://doi.org/10.3758/BF03212971>
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44(2), 325–343. <http://doi.org/10.3758/s13428-011-0146-0>
- Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native Language Influences on Word Recognition in a Second Language: A Megastudy. *Journal of Experimental Psychology: Learning Memory and Cognition*, 34(1), 12–31. <http://doi.org/10.1037/0278-7393.34.1.12>
- Levelt, W. J. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Li, M., Shavelson, R. J., Yin, Y., Wiley, E., Li, M., Shavelson, R. J., ... Wiley, E. (2015). Generalizability Theory. In R. L. Cautin & S. O. Lilienfeld (Eds.), *The Encyclopedia of Clinical Psychology* (pp. 1–19). Hoboken, NJ, USA: John Wiley & Sons, Inc. <http://doi.org/10.1002/9781118625392.wbecp352>
- Libben, & Titone. (2008). The multidetermined nature of idiom processing. *Memory & Cognition*, 36(6), 1103–21. <http://doi.org/10.3758/MC.36.6.1103>
- Liontas, J. (2002). Context and idiom understanding in second languages. *EUROSLA Yearbook*, 2, 155–185. <http://doi.org/10.1075/>

- eurosla.2.11lio
- Liontas, J. (2015a). Developing Idiomatic Competence in the ESOL Classroom: A Pragmatic Account. *TESOL Journal*, 6(4), 621–658. <http://doi.org/10.1002/tesj.230>
- Liontas, J. (2015b). Straight from the Horse's Mouth: Idiomaticity Revisited. In R. R. Heredia & A. B. Ciešlicka (Eds.), *Bilingual Figurative Language Processing* (pp. 301–340). New York: Cambridge University Press. <http://doi.org/10.1017/CBO9781139342100.016>
- Lo Martire, R. (2017). rel: Reliability Coefficients. Retrieved from <https://cran.r-project.org/package=rel>
- Malt, B. C., & Eiter, B. (2004). Even with a green card, you can be put out to pasture and still have to work: non-native intuitions of the transparency of common English idioms. *Memory & Cognition*, 32(6), 896–904. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15673178>
- MathWorks. (2016). Matlab 2016b.
- Meade, G., & Dijkstra, T. (2017). Mechanisms underlying word learning in second language acquisition. In M. Libben, M. Goral, & G. Libben (Eds.), *Bilingualism: A framework for understanding the mental lexicon* (pp. 49–71). John Benjamins Publishing Company.
- Millar, N. (2010). The Processing of Malformed Formulaic Language. *Applied Linguistics*, 32(2), 129–148. <http://doi.org/10.1093/applin/amq035>
- Moore, C. T. (2016). gtheory: Apply Generalizability Theory with R. Retrieved from <https://cran.r-project.org/package=gtheory>
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Retrieved from [http://books.google.nl/books/about/Lexical\\_Phrases\\_and\\_Language\\_Teaching.html?id=VeBluuoZ1wMC&pgis=1](http://books.google.nl/books/about/Lexical_Phrases_and_Language_Teaching.html?id=VeBluuoZ1wMC&pgis=1)
- Neuner, G., Krüger, M., & Grever, U. (1981). *Übungstypologie zum kommunikativen Deutschunterricht*. Berlin: Langenscheidt.
- Nordmann, E., Cleland, A. A., & Bull, R. (2014). Familiarity breeds dissent: Reliability analyses for British-English idioms on measures of familiarity, meaning, literality, and decomposability. *Acta Psychologica*, 149, 87–95. <http://doi.org/10.1016/j.actpsy.2014.03.009>
- Nordmann, E., & Jambazova, A. A. (2017). Normative data for idiomatic expressions. *Behavior Research Methods*, 49, 198–215. <http://doi.org/10.3758/s13428-016-0705-5>
- Nunberg, G. (1979). The non-uniqueness of semantic solutions:

- Polysemy. *Linguistics and Philosophy*, 3(2), 143–184. <http://doi.org/10.1007/BF00126509>
- Oostdijk, N., Reynaert, M., Hoste, V., & Schuurman, I. (2013). The construction of a 500-million-word reference corpus of contemporary written Dutch. In *Essential Speech and Language Technology for Dutch* (pp. 219–247). Springer Verlag.
- Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, 76(3), 241–263. <http://doi.org/10.1037/h0027272>
- Paivio, A. (1986). *Mental representations : a dual coding approach*. Oxford University Press.
- Paivio, A., Yuille, J. C., & Smythe, P. C. (1966). Stimulus and response abstractness, imagery, and meaningfulness, and reported mediators in paired-associate learning. *Canadian Journal of Psychology*, 20(4), 362–77.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (pp. 191–225). London: Longman.
- Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1–2), 8–13. <http://doi.org/10.1016/j.jneumeth.2006.11.017>
- Peters, E. (2012). Learning German formulaic sequences: the effect of two attention-drawing techniques. *The Language Learning Journal*, 40(1), 65–79. <http://doi.org/10.1080/09571736.2012.658224>
- Pollio, H. R., Barlow, J. M., Fine, H. J., & Pollio, M. R. (1977). *Psychology and the Poetics of Growth: Figurative Language in Psychology, Psychotherapy, and Education*. Hillsdale (N.J.): Lawrence Erlbaum Associates.
- Qualtrics. (2005). Qualtrics. Provo, Utah, USA. Retrieved from <http://www.qualtrics.com>
- R Development Core Team. (2008). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Ramonda, K. (2019). The role of encyclopedic world knowledge in semantic. *English Language and Linguistics*, 23(1), 31–53. <http://doi.org/10.1017/S1360674317000284>
- Rietveld, T., & van Hout, R. (1993). *Statistical techniques for the study of language and language behaviour*. Berlin/New York: Mouton de Gruyter.
- Rietveld, T., & van Hout, R. (2018). *Subjective Measuring in Speech-*

- Language Pathology*. Manuscript in preparation.
- Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, 25(5), 762–776. [http://doi.org/10.1162/jocn\\_a\\_00337](http://doi.org/10.1162/jocn_a_00337)
- Sadoski, M. (2005). A Dual Coding View of Vocabulary Learning. *Reading & Writing Quarterly*, 21(3), 221–238. <http://doi.org/10.1080/10573560590949359>
- Sawilowsky, S. S. (2009). New Effect Size Rules of Thumb. *Journal of Modern Applied Statistical Methods*, 8(2), 597–599. <http://doi.org/10.22237/jmasm/1257035100>
- Schepens, J., Dijkstra, T., Grootjen, F., & van Heuven, W. J. B. (2013). Cross-Language Distributions of High Frequency and Phonetically Similar Cognates. *PLoS ONE*, 8(5). <http://doi.org/10.1371/journal.pone.0063006>
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability Theory: a Primer*. Sage Publications.
- Shavelson, R. J., & Webb, N. M. (2006). Generalizability Theory. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of Complementary Methods in Education Research* (pp. 309–322). Mahwah, NJ: Lawrence Erlbaum Associates.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin*, 86(2), 420–428.
- Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, 27(2), 251–272. <http://doi.org/10.1177/0267658310382068>
- Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase “time and again” matters: the role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3), 776–784. <http://doi.org/10.1037/a0022531>
- Siyanova-Chanturia, A., & Martinez, R. (2014). The Idiom Principle Revisited. *Applied Linguistics*, 1–22. <http://doi.org/10.1093/applin/amt054>
- Siyanova-Chanturia, A., & Spina, S. (2015). Investigation of Native Speaker and Second Language Learner Intuition of Collocation Frequency. *Language Learning*, 65(3), 533–562. <http://doi.org/10.1111/lang.12125>
- Siyanova, A., & Schmitt, N. (2008). L2 Learner Production and Processing of Collocation: A Multi-study Perspective. *Canadian Modern*

- Language Review*, 64(3), 429–458. <http://doi.org/10.3138/cmlr.64.3.429>
- Skoufaki, S. (2008). Investigating the Source of Idiom Transparency Intuitions. *Metaphor and Symbol*, 24(1), 20–41.
- Slot Webcommerce BV. (2017). Nederlands Woordenboek - Gezegden. Retrieved from [woorden.org](http://woorden.org)
- Sorhus, H. B. (1977). To hear ourselves - Implications for teaching English as a second language. *English Language Teaching Journal*, 31(3), 211–221.
- Sprenger, S., Levelt, W., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, 54(2), 161–184. <http://doi.org/10.1016/j.jml.2005.11.001>
- Steinel, M. P., Hulstijn, J. H., & Steinel, W. (2007). Second Language Idiom Learning in a Paired-Associate Paradigm: Effects of Direction of Learning, Direction of Testing, Idiom Imageability, and Idiom Transparency. *Studies in Second Language Acquisition*, 29, 449–484.
- Stengers, H., Boers, F., Housen, A., & Eyckmans, J. (2010). Does chunking foster uptake? In S. De Knop, F. Boers, & A. De Rycker (Eds.), *Fostering language teaching efficiency through cognitive linguistics* (pp. 99–118). De Gruyter Mouton.
- Stengers, H., Boers, F., Housen, A., & Eyckmans, J. (2011). Formulaic sequences and L2 oral proficiency: Does the type of target language influence the association? *IRAL - International Review of Applied Linguistics in Language Teaching*, 49(4), 321–343. <http://doi.org/10.1515/iral.2011.017>
- Stengers, H., Deconinck, J., Boers, F., & Eyckmans, J. (2016). Does copying idioms promote their recall? *Computer Assisted Language Learning*, 29(2), 289–301. <http://doi.org/10.1080/09588221.2014.937723>
- Stoett, F. A. (1925). *Nederlandse spreekwoorden, spreekwijzen, uitdrukkingen en gezegden* (4th editio). Zutphen, the Netherlands: W.J. Thieme & Cie.
- Swinney, D., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, 18(5), 523–534. [http://doi.org/10.1016/S0022-5371\(79\)90284-6](http://doi.org/10.1016/S0022-5371(79)90284-6)
- Tabossi, P., Arduino, L., & Fanari, R. (2011). Descriptive norms for 245 Italian idiomatic expressions. *Behavior Research Methods*, 43(1), 110–123. <http://doi.org/10.3758/s13428-010-0018-z>
- Tinsley, H. E. A., & Brown, S. D. (2000). *Handbook of applied multivariate statistics and mathematical modeling*. Academic Press.
- Tinsley, H. E. A., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, 22(4),

- 358–376. <http://doi.org/10.1037/h0076640>
- Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting Bilingual and Monolingual Idiom Processing. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Figurative Language Processing* (pp. 171–207). New York, NY: Cambridge University Press.
- Titone, D., & Connine, C. M. (1994). Descriptive norms for 171 idiomatic expressions: Familiarity, compositionality, predictability, and literality. *Metaphor and Symbolic Activity*, 9(4), 247–270.
- Titone, D., & Libben, M. (2014). Time-dependent effects of decomposability, familiarity and literal plausibility on idiom priming: A cross-modal priming investigation. *The Mental Lexicon*, 9(3), 473–496. <http://doi.org/10.1075/ml.9.3.05tit>
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The Development of Fluency in Advanced Learners of French. *Applied Linguistics*, 17(1), 84–119. <http://doi.org/10.1093/applin/17.1.84>
- Türker, E. (2016). Idiom acquisition by second language learners: the influence of cross-linguistic similarity and context. *The Language Learning Journal*, 1–12. <http://doi.org/10.1080/09571736.2016.1221441>
- Tyler, M. D., Tyler, L., & Burnham, D. K. (2005). The delayed trigger voice key: An improved analogue voice key for psycholinguistic research. *Behavior Research Methods*, 37(1), 139–147. <http://doi.org/10.3758/BF03206408>
- van Ginkel, W., & Dijkstra, T. (2019). The tug of war between an idiom’s figurative and literal meanings: Evidence from native and bilingual speakers. *Bilingualism: Language and Cognition*.
- van Ginkel, W., Hubers, F., Cucchiarini, C., Dijkstra, T., & Strik, H. (2016). Norming Studies for Idiom Processing: Native and Non-Native Benchmarks. In *Paper presented at FLaRN 2016*.
- VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Norwood, NJ: Ablex.
- Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Jens, S. (2004). Invariant Time Course of Priming With and Without Awareness. In C. Kaernbach, E. Schroger, & H. Müller (Eds.), *Psychophysics beyond sensation: Laws and invariants of human cognition* (pp. 271–288). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers. Retrieved from <https://psycnet.apa.org/record/2004-00164-012>
- Warga, M. (2005). “Je serais très merçiable”: Formulaic vs . Creatively Produced Speech in Learners’ Request-Closings. *Canadian Journal of Applied Linguistics*, 8, 67–93.

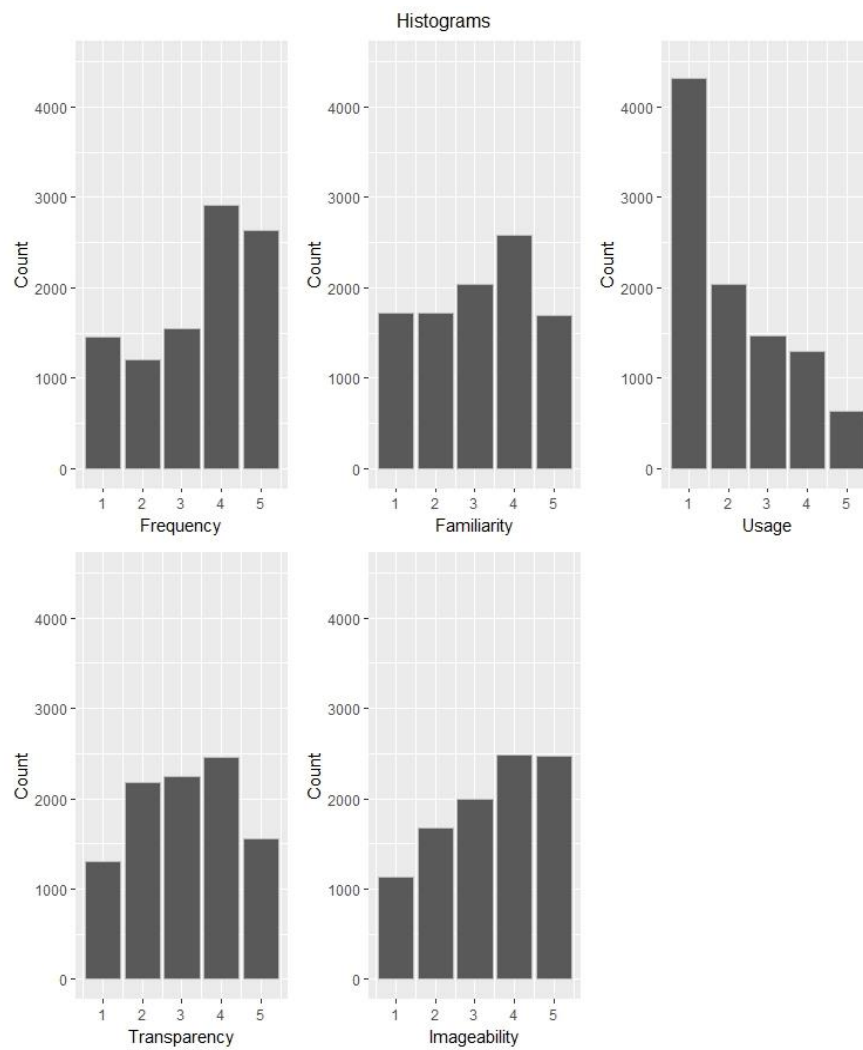
- Wei, L. (2002). The Bilingual Mental Lexicon and Speech Production Process. *Brain and Language*, 81(1-3), 691-707. <http://doi.org/10.1006/BRLN.2001.2557>
- Weinert, R. (1995). The Role of Formulaic Language in Second Language Acquisition: A Review. *Applied Linguistics*, 16(2), 180-205. <http://doi.org/10.1093/applin/16.2.180>
- Wood, D. (2006). Uses and Functions of Formulaic Sequences in Second Language Speech: An Exploration of the Foundations of Fluency. *The Canadian Modern Language Review*, 63(1), 13-33. <http://doi.org/10.3138/cmlr.63.1.13>
- Wood, D. (2010). *Formulaic language and second language speech fluency: background, evidence and classroom applications*. Continuum.
- Wray, A. (2000). Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics*, 21(4), 463-489. <http://doi.org/10.1093/applin/21.4.463>
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press. <http://doi.org/10.1017/CBO9780511519772>
- Wray, A. (2008). *Formulaic Language: Pushing the Boundaries*. *Applied Linguistics* (Vol. 31). Oxford: Oxford University Press.
- Wray, A. (2009). Future directions in formulaic language research. *Journal of Foreign Languages*, 32(6), 2-17.
- Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: an integrated model. *Language & Communication*, 20(1), 1-28. [http://doi.org/10.1016/S0271-5309\(99\)00015-4](http://doi.org/10.1016/S0271-5309(99)00015-4)
- Yeganehjoo, M., & Ngee Thai, Y. (2009). Lexical Access in Production of Idioms by Proficient L2 Learners. *The Southeast Asian Journal of English Language Studies*, 18(1), 87-104.
- Yorio, C. A. (1989). Idiomaticity as an indicator of second language proficiency. In K. Hyltenstam & L. K. Obler (Eds.), *Bilingualism Across the Lifespan: Aspects of Acquisition, Maturity and Loss* (pp. 55-72). Cambridge: Cambridge University Press. <http://doi.org/10.1017/CBO9780511611780.005>
- Zyzik, E. (2011). Second language idiom learning: The effects of lexical knowledge and pedagogical sequencing. *Language Teaching Research*, 15(4), 413-433. <http://doi.org/10.1177/1362168811412025>





## Supplementary materials

### Chapter 2



*Figure S1.* Histograms based on the individual data for the idiom properties Frequency, Familiarity, Usage, Transparency, and Imageability as rated by native speakers

Table S1  
 The Reliability Coefficients per Experimental List and Property, and  
 the Properties of the Experimental Lists (Number of Raters and  
 Number of Items).

<b>Property</b>	<b>List</b>	<b>D-coef</b>	<b>ICC</b>	<b>Cron. <math>\alpha</math></b>	<b>Krip. <math>\alpha</math></b>	<b>#Raters</b>	<b>#Items</b>
Familiarity	1	0.944	0.944	0.9618	0.3541	29	25
Familiarity	2	0.9717	0.9717	0.9787	0.4978	33	25
Familiarity	3	0.9542	0.9542	0.9635	0.3978	30	25
Familiarity	4	0.9686	0.9686	0.9757	0.5039	29	25
Familiarity	5	0.938	0.938	0.9556	0.3149	31	25
Familiarity	6	0.9727	0.9727	0.9789	0.5154	32	25
Familiarity	7	0.9451	0.9451	0.9564	0.3606	29	25
Familiarity	8	0.9709	0.9709	0.9747	0.5325	28	25
Familiarity	9	0.9125	0.9125	0.9282	0.2997	23	25
Familiarity	10	0.9322	0.9322	0.9509	0.4054	19	25
Familiarity	11	0.9087	0.9087	0.9522	0.2834	23	25
Familiarity	12	0.9396	0.9396	0.958	0.4109	21	25
Familiarity	13	0.9102	0.9102	0.9353	0.2913	23	25
Familiarity	14	0.9118	0.9118	0.9329	0.3267	20	25
Familiarity	15	0.9636	0.9636	0.9709	0.5579	20	25
Frequency	1	0.9504	0.9504	0.9641	0.3849	29	25
Frequency	2	0.966	0.966	0.9753	0.4503	33	25
Frequency	3	0.9383	0.9383	0.9519	0.3243	30	25
Frequency	4	0.974	0.974	0.9777	0.5531	29	25
Frequency	5	0.9402	0.9402	0.9522	0.325	31	25
Frequency	6	0.9755	0.9755	0.9795	0.5434	32	25
Frequency	7	0.9492	0.9492	0.958	0.3802	29	25
Frequency	8	0.9633	0.9633	0.9704	0.4724	28	25
Frequency	9	0.9215	0.9215	0.9426	0.3238	23	25
Frequency	10	0.9418	0.9418	0.9565	0.4465	19	25
Frequency	11	0.9174	0.9174	0.9446	0.3101	23	25
Frequency	12	0.929	0.929	0.9501	0.3691	21	25
Frequency	13	0.927	0.927	0.9414	0.3431	23	25
Frequency	14	0.8942	0.8942	0.9245	0.2814	20	25
Frequency	15	0.9579	0.9579	0.9664	0.5204	20	25

Usage	1	0.9496	0.9496	0.9619	0.3811	29	25
Usage	2	0.9642	0.9642	0.9757	0.4369	33	25
Usage	3	0.9264	0.9264	0.9496	0.2819	30	25
Usage	4	0.9644	0.9644	0.975	0.4701	29	25
Usage	5	0.9551	0.9551	0.9658	0.3948	31	25
Usage	6	0.9663	0.9663	0.9746	0.4604	32	25
Usage	7	0.9545	0.9545	0.9644	0.4078	29	25
Usage	8	0.9528	0.9528	0.9649	0.4063	28	25
Usage	9	0.8808	0.8808	0.9143	0.2284	23	25
Usage	10	0.9361	0.9361	0.9472	0.4226	19	25
Usage	11	0.8654	0.8654	0.9189	0.2005	23	25
Usage	12	0.8963	0.8963	0.9289	0.2756	21	25
Usage	13	0.9412	0.9412	0.9554	0.3969	23	25
Usage	14	0.8883	0.8883	0.9297	0.2664	20	25
Usage	15	0.9449	0.9449	0.9586	0.448	20	25
Transparency	1	0.9081	0.9081	0.9347	0.2409	29	25
Transparency	2	0.8417	0.8417	0.8889	0.1267	33	25
Transparency	3	0.8933	0.8933	0.9275	0.2046	30	25
Transparency	4	0.8008	0.8008	0.8601	0.1086	29	25
Transparency	5	0.8952	0.8952	0.9171	0.205	31	25
Transparency	6	0.9119	0.9119	0.9465	0.2298	32	25
Transparency	7	0.8728	0.8728	0.9108	0.178	29	25
Transparency	8	0.865	0.865	0.8934	0.1749	28	25
Transparency	9	0.8916	0.8916	0.9119	0.2512	23	25
Transparency	10	0.8971	0.8971	0.9237	0.2994	19	25
Transparency	11	0.7705	0.7705	0.8341	0.1123	23	25
Transparency	12	0.839	0.839	0.8943	0.181	21	25
Transparency	13	0.8799	0.8799	0.9241	0.2246	23	25
Transparency	14	0.824	0.824	0.8636	0.1754	20	25
Transparency	15	0.9049	0.9049	0.9413	0.3046	20	25
Imageability	1	0.9321	0.9321	0.9479	0.3088	29	25
Imageability	2	0.9071	0.9071	0.92	0.2189	33	25
Imageability	3	0.9096	0.9096	0.9308	0.2392	30	25
Imageability	4	0.8629	0.8629	0.8958	0.1664	29	25
Imageability	5	0.9041	0.9041	0.9155	0.2238	31	25
Imageability	6	0.8895	0.8895	0.9061	0.1914	32	25

Imageability	7	0.9342	0.9342	0.9465	0.3169	29	25
Imageability	8	0.922	0.9224	0.9404	0.2929	27	25
Imageability	9	0.8415	0.8415	0.8912	0.1717	23	25
Imageability	10	0.8479	0.8479	0.873	0.2143	19	25
Imageability	11	0.7376	0.7376	0.8204	0.0921	23	25
Imageability	12	0.8255	0.8255	0.8726	0.1684	21	25
Imageability	13	0.8982	0.8982	0.9227	0.2635	23	25
Imageability	14	0.8107	0.8107	0.8575	0.1612	20	25
Imageability	15	0.9311	0.9311	0.9477	0.3895	20	25

Table S2

Variance Inflation Factors (VIFs) for the final regression analysis.

<b>Fixed effects</b>	<b>VIF</b>
Familiarity	2.31
Transparency	1.14
Imageability	1.62
Frequency	1.90
Usage	1.49
Objective idiom frequency	1.02
Familiarity x Transparency	1.37
Familiarity x Imageability	1.56

## Chapter 3

Table S3

Dutch idiomatic expressions used in the rating study reported on in Chapter 3

#	Idiomatic expression	Meaning
1	een wassen neus	niet van belang
2	het op zijn heupen krijgen	plotseling fanatiek bezig gaan
3	in het stof bijten	verliezen
4	iemand tegen de schenen schoppen	vervelend doen tegen iemand
5	veel noten op zijn zang hebben	verwaand zijn
6	iemand de wet voorschrijven	bepalen wat iemand moet doen
7	de koe bij de hoorns vatten	een flinke klus aanpakken
8	iemand een koekje van eigen deeg geven	iemand behandelen zoals hij anderen behandelt
9	zich in de nesten werken	problemen krijgen door eigen gedrag
10	het bij het verkeerde eind hebben	ongelijk hebben
11	iets onder de pet houden	iets geheim houden
12	van de oude stempel zijn	iets op een oude manier doen
13	ergens een potje van maken	iets doen mislukken
14	iemand in de maling nemen	een grap met iemand uithalen
15	iets niet aan de grote klok hangen	iets niet algemeen bekend maken
16	iets uit je duim zuigen	iets verzinnen
17	voor aap staan	zichzelf belachelijk maken
18	aan de slag gaan met iets	beginnen met iets
19	de sterren van de hemel spelen	erg goed spelen
20	iemand een hart onder de riem steken	iemand moed inspreken
21	iets komt voor de bakker	het komt in orde
22	iets over het hoofd zien	iets niet opmerken
23	met een schone lei beginnen	opnieuw beginnen zonder last van het verleden
24	van de hak op de tak springen	van het ene op het andere overgaan zonder een verband te leggen

25	zich uit de naad werken	heel hard werken
26	het schip ingaan	verliezen
27	duit in het zakje doen	een bijdrage leveren
28	boter bij de vis	contant betalen
29	een klein hartje hebben	gauw bang zijn
30	ergens mee voor de draad komen	iets vertellen
31	op zijn achterste benen staan	je ergens tegen verzetten
32	iets op je lever hebben	willen zeggen wat je dwars zit
33	aan een zijden draadje hangen	weinig kans op succes hebben
34	iets uit de losse pols doen	iets zonder voorbereiding uitvoeren
35	ergens mee in de maag zitten	niet goed weten wat te doen met iets
36	iets van tafel vegen	iets zonder overleg afwijzen
37	een hoofd als een boei krijgen	erg blozen
38	iemand van haver tot gort kennen	iemand erg goed kennen
39	van de baan zijn	niet meer doorgaan
40	aan het roer staan	de leiding hebben
41	het ver schoppen	succesvol zijn
42	iemand wakker schudden	iemand krachtig aan iets herinneren
43	iets stuit iemand tegen de borst	afkeer hebben van iets
44	op zijn centen zitten	gierig zijn
45	aan de bak komen	werk vinden
46	de dans ontspringen	aan iets vervelends ontkomen
47	het klopt als een bus	het klopt helemaal
48	iemand voor geen cent vertrouwen	iemand helemaal niet vertrouwen
49	iets over een andere boeg gooien	iets op een andere manier aanpakken
50	met de gebakken peren zitten	ergens voor moeten opdraaien
51	oogje in het zeil houden	toezicht houden op iets
52	voet bij stuk houden	niet toegeven
53	met de hand op het hart	oprecht
54	iemand om zeep helpen	iemand vermoorden
55	op de kast zitten	boos zijn

56	iemand iets in de maag splitsen	iemand iets dwingen te doen
57	iemand in de kaart spelen	iemand onbedoeld helpen
58	iets soldaat maken	iets opeten
59	een slag om de arm houden	iets onder voorbehoud afspreken
60	lange tenen hebben	snel beledigd zijn
61	één lijn trekken	dezelfde aanpak gebruiken
62	iets op poten zetten	iets nieuws opstarten
63	een vinger aan de pols houden	steeds controleren
64	iemand iets op de mouw spelden	iemand iets wijsmaken
65	druk van de ketel halen	enige rust brengen
66	iemand in de steek laten	iemand op een kritiek moment niet helpen
67	in goede aarde vallen	gewaardeerd worden
68	iemand iets in de schoenen schuiven	iemand onterecht beschuldigen
69	het niet breed hebben	niet veel geld hebben
70	iemand op de huid zitten	iemand continu, op een vervelende manier, controleren
71	iets op zijn duimpje kennen	iets zeer goed kennen
72	met de rug tegen de muur staan	geen keuze hebben
73	zich in bochten wringen	op alle mogelijke wijzen iets proberen
74	buiten westen zijn	bewusteloos zijn
75	de moed in de schoenen zakken	alle moed verliezen
76	iemand het vuur na aan de schenen leggen	iemand onder druk zetten
77	iets op prijs stellen	iets waarderen
78	in de soep lopen	mislukken
79	niet door de beugel kunnen	de norm overschrijden
80	veel in zijn mars hebben	veel kunnen of weten
81	uit de school klappen	geheimen vertellen
82	goed beslagen ten ijs komen	goed voorbereid zijn
83	iets met de Franse slag doen	iets haastig en slordig doen
84	de boeken sluiten	bankroet gaan
85	veel voeten in de aarde hebben	veel moeite kosten
86	tussen twee vuren zitten	van twee kanten bedreigd worden

87	koek en ei zijn	goede vrienden zijn
88	als paddenstoelen uit de grond schieten	snel en in grote massa tevoorschijn komen
89	ergens zijn neus voor optrekken	zich te goed vinden voor iets
90	iemand tot in de wolken verheffen	iemand uitbundig prijzen
91	de langste adem hebben	iets het langst volhouden
92	iemand een loer draaien	iemand belazeren
93	de vruchten plukken van iets	voordeel hebben van iets
94	het kind van de rekening zijn	een slachtoffer zijn
95	iets op de lange baan schuiven	iets uitstellen
96	iemand aan de tand voelen	iemand streng ondervragen
97	hard van stapel lopen	zich overhaasten
98	iemand om de tuin leiden	iemand misleiden
99	iets onder de knie hebben	iets goed kunnen
100	iemand onder de duim hebben	iemand in bedwang houden
101	voor iemand door het vuur gaan	voor iemand alles overhebben
102	achter het net vissen	zijn kans missen
103	door het lint gaan	je woede niet meer kunnen beheersen
104	iemand een rad voor de ogen draaien	iemand misleiden
105	iets op je buik kunnen schrijven	niet krijgen wat je hebben wil
106	iets van de bovenste plank	iets van de beste kwaliteit
107	naar de haaien gaan	kapot gaan
108	van streek zijn	ontdaan zijn
109	voor spek en bonen	zonder mee te tellen
110	het onderste uit de kan halen	alles willen hebben

---



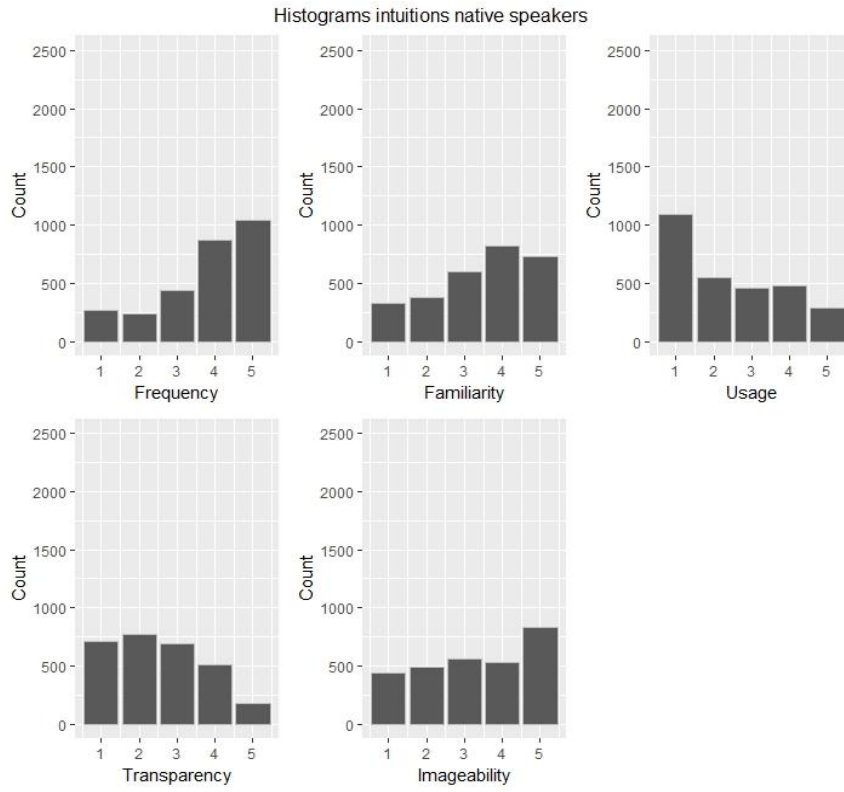
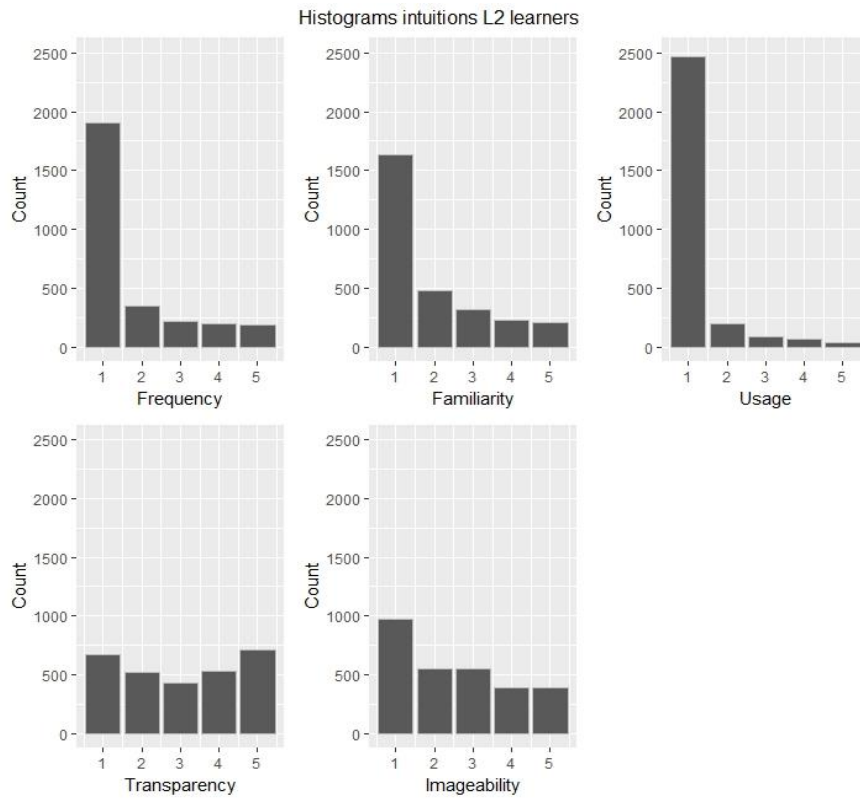


Figure S2. Histograms for the idiom properties Frequency, Familiarity, Usage, Transparency, and Imageability as rated by the native speakers.



*Figure S3.* Histograms for the idiom properties Frequency, Familiarity, Usage, Transparency, and Imageability as rated by the L2 learners.

## Chapter 4

Table S4

Dutch idiomatic expressions included in the CALL-based learning experiment

#	Idiomatic expression	Meaning
1	niet goed uit de verf komen	niet goed bij anderen overkomen
2	aan de grond zitten	in slechte omstandigheden verkeren
3	een vinger in de pap hebben	invloed hebben op iets
4	het op zijn heupen krijgen	plotseling fanatiek bezig gaan
5	hoog van de toren blazen	opscheppen
6	iemand iets in de maag splitsen	iemand iets dwingen te doen
7	iemand iets op de mouw spelden	iemand iets wijsmaken
8	iets onder de knie hebben	iets goed kunnen
9	iets soldaat maken	iets opeten
10	iets uit de doeken doen	iets uitleggen
11	lange tenen hebben	snel beledigd zijn
12	naast zijn schoenen lopen	zich arrogant gedragen
13	op de tocht staan	in een bedreigde positie komen
14	tegen de lamp lopen	betrapt worden
15	tegen het plafond zitten	niet meer kunnen bereiken
16	boter bij de vis	contant betalen
17	bij iemand een wit voetje halen	bij iemand in de gunst proberen te komen
18	de kat op het spek binden	iemand in verleiding brengen
19	een klein hartje hebben	gauw bang zijn
20	een slag om de arm houden	iets onder voorbehoud afspreken
21	een wassen neus	niet van belang
22	hek van de dam	geen belemmeringen meer hebben
23	iemand in de kaart spelen	iemand onbedoeld helpen
24	iemand om zeep helpen	iemand vermoorden
25	koek en ei zijn	goede vrienden zijn
26	met zijn neus in de boter vallen	in een gunstige situatie terechtkomen
27	op de fles gaan	failliet gaan
28	op een laag pitje staan	minder aandacht krijgen

29	op zijn strepen staan	zijn eigen mening aanhouden
30	veel voeten in de aarde hebben	veel moeite kosten
31	voor spek en bonen	zonder mee te tellen
32	aan een zijden draadje hangen	weinig kans op succes hebben
33	aan het roer staan	de leiding hebben
34	als paddenstoelen uit de grond schieten	snel en in grote massa tevoorschijn komen
35	de koe bij de hoorns vatten	een flinke klus aanpakken
36	de mouwen opstropen	aan het werk gaan
37	één lijn trekken	dezelfde aanpak gebruiken
38	ergens mee in de maag zitten	niet goed weten wat te doen met iets
39	ergens zijn neus voor optrekken	zich te goed vinden voor iets
40	het niet breed hebben	niet veel geld hebben
41	aan de tand voelen	iemand streng ondervragen
42	iemand de wet voorschrijven	bepalen wat iemand moet doen
43	iemand iets in de schoenen schuiven	iemand onterecht beschuldigen
44	iemand op de huid zitten	iemand continu, op een vervelende manier, controleren
45	iemand tegen de schenen schoppen	vervelend doen tegen iemand
46	iemand wakker schudden	iemand krachtig aan iets herinneren
47	iemand tot in de wolken verheffen	iemand uitbundig prijzen
48	iets op poten zetten	iets nieuws opstarten
49	iets uit de losse pols doen	iets zonder voorbereiding uitvoeren
50	iets uit je duim zuigen	iets verzinnen
51	iets van tafel vegen	iets zonder overleg afwijzen
52	met de rug tegen de muur staan	geen keuze hebben
53	op het puntje van de tong liggen	niet op een woord kunnen komen
54	op zijn achterste benen staan	je ergens tegen verzetten
55	op zijn centen zitten	gierig zijn
56	tussen twee vuren zitten	van twee kanten bedreigd worden
57	voor aap staan	zichzelf belachelijk maken
58	voor iemand door het vuur gaan	voor iemand alles overhebben
59	geen hart in het lijf hebben	geen medelijden kennen
60	het hoofd koel houden	rustig blijven

---

het op zijn heupen krijgen

plotseling fanatiek bezig gaan

resterende tijd: 


*Figure S4.* Screenshot of the PAL paradigm exercise. The upper phrase is the Dutch idiomatic expression, and the phrase below is the corresponding meaning. Participants were instructed to carefully read the idiom-meaning pair. No explicit task was formulated.

Vraag 1 1/30

Vul het ontbrekende woord in:

Het verlegen meisje durfde uiteindelijk het verhaal uit de ...  
te doen

Schrijf je antwoord op:

resterende tijd: 

[Volgende vraag](#)

Figure S5. Screenshot of the gap-fill exercise.

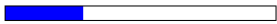
Vraag 1 1/30

Maak de zin af door de juiste uitdrukking te kiezen:

Zij reageerde heel verontwaardigd, ze

- heeft lange tenen
- zit tegen het plafond
- viel met haar neus in de boter

Schrijf je antwoord op:

resterende tijd: 

[Volgende vraag](#)

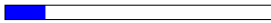
Figure S6. Screenshot of the sentence completion exercise.

Vraag 1 1/30

Kies de juiste uitdrukking:  
zich te goed vinden voor iets

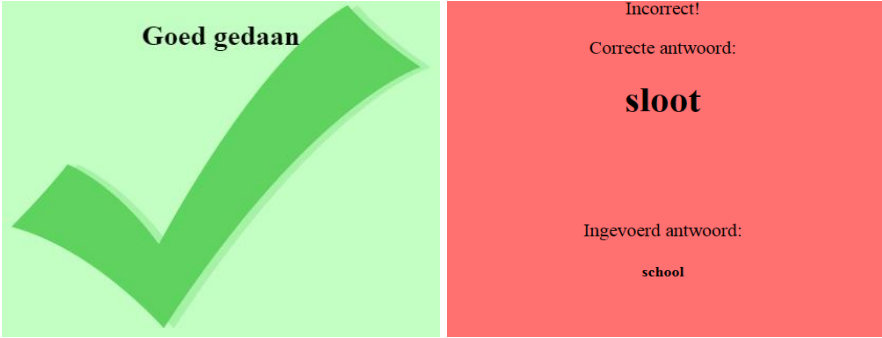
- het op zijn heupen krijgen
- ergens zijn neus voor optrekken
- tegen het plafond zitten

Schrijf je antwoord op:

resterende tijd: 

[Volgende vraag](#)

Figure S7. Screenshot of the idiom selection exercise.



The image shows two side-by-side panels of feedback. The left panel has a light green background and features a large green checkmark with the text "Goed gedaan" (Well done) above it. The right panel has a light red background and displays the text "Incorrect!" at the top, followed by "Correcte antwoord:" (Correct answer:) and the word "sloot" in a large, bold font. Below that, it says "Ingevoerd antwoord:" (Entered answer:) and the word "school" in a smaller font.

Figure S8. Screenshot of the corrective feedback participants received directly after providing the answer. The left panel: Feedback in case of a correct answer. The right panel: The feedback in case of an incorrect answer.



Figure S9. Screenshot of the reading aloud manipulation.



Figure S10. Screenshot of the reading silently manipulation.



## Chapters 5 and 6

Table S5  
Idiomatic expressions used in the experiments reported on in Chapters 5 and 6

#	Idiomatic expression	Meaning	COR	REL	UNREL
1	aan de grond zitten	in slechte omstandigheden verkeren	grond	lucht	droom
2	bij iemand een wit voetje halen	bij iemand in de gunst proberen te komen	voetje	teentje	bosje
3	boter bij de vis	contant betalen	vis	zee	grap
4	de kat op het spek binden	iemand in verleiding brengen	spek	vet	gaas
5	een klein hartje hebben	gauw bang zijn	hartje	kusje	hoekje
6	een slag om de arm houden	iets onder voorbehoud afspreken	arm	pols	rust
7	een vinger in de pap hebben	invloed hebben op iets	pap	melk	krant
8	een wassen neus	niet van belang	neus	oog	trein
9	hek van de dam	geen belemmeringen meer hebben	dam	bever	dolk
10	het op zijn heupen krijgen	plotseling fanatiek bezig gaan	heupen	billen	bedden
11	hoog van de toren blazen	opscheppen	toren	klok	leden
12	iemand iets in de maag splitsen	iemand iets dwingen te doen	maag	buik	huur
13	iemand iets op de mouw spelden	iemand iets wijsmaken	mouw	trui	slee
14	iemand in de kaart spelen	iemand onbedoeld helpen	kaart	brief	helft
15	iemand om zeep helpen	iemand vermoorden	zeep	sop	munt

Table S5 continued  
 Idiomatic expressions used in the experiments reported on in Chapters 5 and 6

#	Idiomatic expression	Meaning	COR	REL	UNREL
16	iets onder de knie hebben	iets goed kunnen	knie	schijf	villa
17	iets soldaat maken	zich arrogant gedragen	soldaat	majoor	getuige
18	iets uit de doeken doen	iets uitleggen	doeken	lakens	herten
19	koek en ei zijn	goede vrienden zijn	ei	ham	zaal
20	lange tenen hebben	snel beledigd zijn	tenen	nagels	apen
21	met zijn neus in de boter vallen	in een gunstige situatie terecht komen	boter	pan	klem
22	naast zijn schoenen lopen	zich arrogant gedragen	schoenen	benen	partners
23	op de fles gaan	failliet gaan	fles	drank	tuin
24	op de tocht staan	in een bedreigde positie komen	tocht	kou	lente
25	op een laag pitje staan	minder aandacht krijgen	pitje	vlammetje	kiertje
26	op zijn strepen staan	zijn eigen mening aanhouden	strepen	lijnen	messen
27	tegen de lamp lopen	betraht worden	lamp	warmte	helm
28	tegen het plafond zitten	niet meer kunnen bereiken	plafond	dak	formuis
29	veel voeten in de aarde hebben	veel moeite kosten	aarde	hemel	onzin
30	voor spek en bonen	zonder mee te tellen	bonen	erwtten	beren

## Nederlandse samenvatting

Taal bestaat voor een groot deel uit vaste combinaties van woorden. Voorbeelden zijn *sterke drank*, *tegen de lamp lopen*, *rekenen op*, en *vraag en aanbod*. Deze woorden worden zo vaak in deze combinaties gebruikt, dat het gek klinkt om wat te veranderen aan die combinaties. Neem bijvoorbeeld de woordcombinatie *sterke drank*. Als je het woord *sterke* zou vervangen door een synoniem, *krachtige*, dan leidt dit tot een vreemde of atypische combinatie van woorden. In het geval van de combinatie *vraag en aanbod* leidt het verwisselen van de woorden *vraag* en *aanbod* tot zo'n atypische combinatie. Een moedertaalspreker van het Nederlands zal zelden de combinatie *krachtige drank* en *aanbod en vraag* gebruiken, terwijl de veranderingen niet leiden tot een andere betekenis. *Krachtige drank* betekent hetzelfde als *sterke drank* en *aanbod en vraag* betekent hetzelfde als *vraag en aanbod*.

Bij woordcombinaties zoals *tegen de lamp lopen* leiden dit soort veranderingen echter wel tot een verschil in betekenis. Deze combinaties noemen we ook wel *idiomatische uitdrukkingen* of *idiomen*. Deze uitdrukkingen hebben een betekenis die vaak los staat van de betekenis van de individuele woorden: een figuratieve betekenis. Als uitdrukking betekent *tegen de lamp lopen* bijvoorbeeld niet dat iemand letterlijk tegen een lamp aanloopt, maar dat deze persoon betrappt wordt. Niets in de woorden *tegen*, *lamp* en *lopen* maakt duidelijk dat het hier om de betekenis 'betrappt worden' gaat. Wanneer in deze uitdrukking het woord *lamp* vervangen wordt door het woord *kamerlamp*, een specifieke lamp, dan verliest deze woordcombinatie zijn figuratieve betekenis. *Tegen de kamerlamp lopen* kan alleen nog maar betekenen dat iemand letterlijk tegen een lamp aanloopt.

Idiomatische uitdrukkingen en andere vaste woordcombinaties vormen een belangrijk onderdeel van taal. Moedertaalsprekers zijn over het algemeen goed in staat om idiomatische uitdrukkingen te begrijpen en te gebruiken. In veel gevallen hebben moedertaalsprekers niet eens in de gaten dat iets een uitdrukking is. De woordcombinatie *in de gaten hebben* uit de vorige zin is een voorbeeld van zo'n onopvallende idiomatische uitdrukking. Nederlanders weten precies wat er hier bedoeld wordt en zullen die combinatie van woorden nooit letterlijk

interpreteren. Voor niet-moedertaalsprekers of tweede-taalleerders (T2-leerders) zijn uitdrukkingen vaak echter lastig te leren. Zelfs als ze de taal al erg goed spreken kunnen ze hier nog moeite mee hebben. Ze gebruiken een stuk minder idiomatische uitdrukkingen dan moedertaalsprekers (Kecskes, 2007, 2015).

Toch is het belangrijk dat deze T2-leerders idiomatische uitdrukkingen onder de knie krijgen. Een groot gedeelte van ons taalgebruik is namelijk figuratief van aard (zie bv. Pawley & Syder, 1983). Moedertaalsprekers gebruiken idiomatische uitdrukkingen en andere vaste combinaties aan de lopende band. Onderzoek heeft uitgewezen dat T2-leerders als vloeiender en taalvaardiger overkomen als ze idiomen en andere vaste woordcombinaties gebruiken (Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006). Ook hechten T2-leerders zelf vaak ook veel belang aan het leren van uitdrukkingen (Liontas, 2015b).

### **Eigenschappen van uitdrukkingen**

Uitdrukkingen zijn er in allerlei soorten en maten. Ze kunnen in meerdere eigenschappen van elkaar verschillen (Cieślicka, 2015; Titone et al., 2015). Afhankelijk van hun eigenschappen zijn uitdrukkingen bijvoorbeeld moeilijker of gemakkelijker te leren of te begrijpen voor T2-leerders en moedertaalsprekers. Eerder onderzoek heeft aangetoond dat deze eigenschappen een belangrijke rol spelen bij het begrijpen, gebruiken en leren van idiomatische uitdrukkingen. De onderstaande Nederlandse uitdrukkingen zijn allemaal afkomstig uit een grootschalige database van 374 uitdrukkingen die we in het kader van ons onderzoeksproject hebben samengesteld (Hubers et al., 2018).

Een van deze eigenschappen is *frequentie of bekendheid*. Sommige uitdrukkingen komen nu eenmaal vaker voor in het dagelijks leven dan andere. Ze verschillen in hun frequentie van voorkomen. Dit heeft natuurlijk ook gevolgen voor het wel of niet weten wat de uitdrukking betekent. Van hoogfrequente uitdrukkingen zullen mensen vaak weten wat ze betekenen, terwijl dit voor laagfrequente uitdrukkingen minder vaak het geval is. Een voorbeeld van een hoogfrequente en bekende uitdrukking in het Nederlands is *onder de knie krijgen* ('in staat zijn om iets te doen'). Deze uitdrukking wordt dikwijls gebruikt in het dagelijks leven. Een voorbeeld van een relatief onbekende en laagfrequente

uitdrukking is *varkensvlees onder de armen hebben* ('lui zijn'). Maar weinig mensen weten wat deze uitdrukking betekent.

Andere eigenschappen waarin uitdrukkingen van elkaar kunnen verschillen zijn *letterlijkheid* en *transparantie*. *Letterlijkheid*, of *letterlijke plausibiliteit*, heeft te maken met de mate waarin je een uitdrukking letterlijk kunt gebruiken (Titone & Libben, 2014). Sommige uitdrukkingen kunnen gemakkelijk letterlijk gebruikt worden. De uitdrukking *tegen de lamp lopen* ('betragt worden') is hier een voorbeeld van. Voor andere uitdrukkingen, zoals *het loodje leggen* ('doodgaan'), is dit een stuk moeilijker. *Transparantie* verwijst naar de mate waarin je op basis van de afzonderlijke woorden de figuratieve betekenis kunt herleiden (Cieślicka, 2015). Voorbeelden van transparante uitdrukkingen zijn *aan het roer staan* ('de leiding hebben') en *onder het mes gaan* ('geopereerd worden'). Als je geen idee hebt wat deze uitdrukkingen betekenen, ben je op basis van de individuele woorden waarschijnlijk toch in staat om de corresponderende figuratieve betekenis te achterhalen. Voor uitdrukkingen zoals *het loodje leggen* en *iets soldaat maken* ('iets opeten of opdrinken') is dit een stuk moeilijker. De individuele woorden geven weinig informatie over de figuratieve betekenis. We noemen deze uitdrukkingen daarom niet-transparant.

Een andere dimensie waarin uitdrukkingen van elkaar kunnen verschillen is *visualiseerbaarheid*. Deze eigenschap wordt vaak gedefinieerd als de mate waarin je een beeld kunt vormen bij de uitdrukking (Cacciari & Glucksberg, 1995; Steinel et al., 2007). Voorbeelden van visualiseerbare uitdrukkingen zijn *in het water vallen* ('mislukken') en *iemand de rug toekeren* ('zich van iemand afwenden'). Deze uitdrukkingen roepen een duidelijk beeld op. Dit is anders voor niet-visualiseerbare uitdrukkingen zoals *iemand een loer draaien* ('iemand belazeren') en *iets met de Franse slag doen* ('iets haastig en slordig doen'). De beelden die uitdrukkingen oproepen kunnen betrekking hebben op zowel de figuratieve als de letterlijke interpretatie van de uitdrukking. Onderzoek heeft echter aangetoond dat moedertaalsprekers zich vaker een beeld vormen van de letterlijke betekenis dan van de figuurlijke betekenis (Cacciari & Glucksberg, 1995).

De bovenstaande eigenschappen spelen een rol bij zowel moedertaalsprekers als bij T2-leerders. Een eigenschap die alleen van

belang is voor T2-leerders is *gelijkenis met de moedertaal*. Deze eigenschap verwijst naar de mate van overlap of gelijkenis in vorm en betekenis tussen uitdrukkingen in de tweede taal en de moedertaal van T2-leerders. Sommige uitdrukkingen zijn identiek in vorm en betekenis in de twee talen. De Nederlandse uitdrukking *de eerste viool spelen* bestaat bijvoorbeeld ook als een directe vertaling in het Engels (*to play the first fiddle*) en het Duits (*die Erste Geige spielen*) en dragen dezelfde betekenis ('de belangrijkste rol vervullen'). Er zijn ook uitdrukkingen die dezelfde betekenis hebben in twee talen, en waarin veel, maar niet alle woorden hetzelfde zijn. *De koe bij de hoorns vatten* ('een flinke klus aanpakken') is hier een voorbeeld van. De Engelse en Duitse varianten van deze uitdrukking zijn *to take the bull by the horns* en *den Stier bei den Hörnen fassen*. Hoewel de meeste woorden uit deze uitdrukkingen overeenkomen met die in de Nederlandse uitdrukkingen, gaat het in het Engels en het Duits niet over een koe, maar over een stier. In deze voorbeelden is er in bepaalde mate sprake van vormoverlap. Er zijn echter ook uitdrukkingen in meerdere talen die dezelfde betekenis uitdrukken, maar met heel andere woorden. De Nederlandse uitdrukking *water naar de zee dragen* ('nutteloos werk doen') is hier een voorbeeld van. In het Engels spreekt men van kolen naar Newcastle dragen (*to carry coals to Newcastle*) en in het Duits van uilen naar Athene dragen (*Eulen nach Athen tragen*). Deze uitdrukkingen hebben allemaal dezelfde betekenis, maar de gebruikte woorden zijn geen vertalingen. Er zijn ook uitdrukkingen in de ene taal waarvan er geen vergelijkbare uitdrukkingen bestaan in de andere taal met dezelfde betekenis. De Nederlandse uitdrukking *goed uit de verf komen* ('goed bij anderen overkomen') heeft bijvoorbeeld geen exact equivalent in het Engels en het Duits. Er is in deze talen geen uitdrukking die dezelfde betekenis draagt. Als de moedertaal en de tweede taal van een leerder dicht bij elkaar staan en er zijn veel overeenkomsten wat betreft hun cultuur (zoals in het Nederlands en Duits), dan is de kans groter dat er meerdere uitdrukkingen met een grote mate van gelijkenis bestaan, vergeleken met talen en culturen die verder van elkaar af staan, zoals het Nederlands en het Chinees.

### **Doel van dit proefschrift**

Op dit moment is het nog onduidelijk hoe T2-leerders omgaan met idiomatische uitdrukkingen in hun tweede taal en in hoeverre ze vergelijkbaar zijn met moedertaalsprekers van die taal. Hoe verwerken T2-leerders deze uitdrukkingen? En wat is het effect van ervaring en systematische oefening met uitdrukkingen op de verwerking en verwerving van uitdrukkingen in een tweede taal? Hoewel deze vragen sterk aan elkaar gerelateerd zijn, heeft eerder onderzoek ze vooral apart bestudeerd. Dit proefschrift onderzoekt de bovenstaande vragen in samenhang door idiomatische uitdrukkingen te bestuderen in de eerste (T1) en tweede taal (T2). Het doel van het proefschrift is om aspecten van idioomverwerking door T2-leerders te vergelijken met die door moedertaalsprekers. Meer specifiek stelden we drie onderzoeksvragen centraal.

- (1) In hoeverre zijn T2-leerders vergelijkbaar met moedertaalsprekers als het gaat om de kennis, representatie en verwerking van idiomatische uitdrukkingen?
- (2) In hoeverre gaan T2-leerders meer op moedertaalsprekers lijken wat betreft de bovenstaande aspecten door uitdrukkingen systematisch aan te leren?
- (3) Wat is het effect van verschillende eigenschappen van uitdrukkingen op het leren en verwerken van idiomatische uitdrukkingen door T2-leerders in vergelijking met moedertaalsprekers?

Antwoorden op deze onderzoeksvragen kunnen ons inzicht geven in de onderliggende mechanismen die T2-leerders en moedertaalsprekers gebruiken om uitdrukkingen te leren, begrijpen en gebruiken en hoe gericht oefenen deze mechanismen beïnvloedt. Op basis hiervan kunnen we zien of T2-leerders fundamenteel anders omgaan met idiomatische uitdrukkingen dan moedertaalsprekers.

De bovenstaande onderzoeksvragen komen in de hoofdstukken 2 tot en met 6 aan bod. In Hoofdstuk 7 worden alle antwoorden op deze vragen geïntegreerd en presenteren we de conclusies.

## Hoofdstuk 2. Normatieve data over Nederlandse idiomatische uitdrukkingen van moedertaalsprekers

Om te onderzoeken hoe T2-leerders omgaan met idiomatische uitdrukkingen is het belangrijk om ze te vergelijken met moedertaalsprekers. Daarom onderzochten we in Hoofdstuk 2 eerst de idioomkennis van moedertaalsprekers van het Nederlands en hun intuïties over de eigenschappen van deze uitdrukkingen.

In een grootschalige studie hebben we 390 moedertaalsprekers van het Nederlands gevraagd om idiomatische uitdrukkingen te beoordelen op verschillende eigenschappen (*frequentie*, *bekendheid*, *gebruik*, *transparantie*, en *visualiseerbaarheid*). Daarnaast vroegen we deze proefpersonen om aan te geven wat de uitdrukkingen betekenden in een meerkeuzevraag en een open vraag. De eigenschappen van de uitdrukkingen werden beoordeeld op 5-puntsschalen. In totaal hebben we data verzameld voor 374 Nederlandse idiomatische uitdrukkingen. Deze database is de eerste grootschalige database voor het Nederlands (Hubers et al., 2018; toegankelijk via <https://doi.org/10.17026/dans-zjx-hnsk>).

Met dit onderzoek wilden we erachter komen welke uitdrukkingen bekend zijn bij moedertaalsprekers, hoe betrouwbaar hun intuïties over eigenschappen van deze uitdrukkingen zijn en in hoeverre deze eigenschappen hun kennis van idiomatische uitdrukkingen beïnvloeden. Hiermee konden we op een weloverwogen manier idiomatische uitdrukkingen te selecteren voor gebruik in latere studies. Om de betrouwbaarheid van de intuïties te bepalen, gebruikten we een maat die nieuw was voor dit vakgebied, de D-coëfficiënt.

Uit de resultaten bleek dat de idioomkennis van moedertaalsprekers van het Nederlands heel goed was. Daarnaast vonden we dat de intuïties over eigenschappen van de uitdrukkingen erg betrouwbaar waren, dat wil zeggen vergelijkbaar voor al onze proefpersonen. Steeds dezelfde uitdrukkingen werden bijvoorbeeld als frequent of transparant beoordeeld door het merendeel van de proefpersonen. De betrouwbaarheid van de intuïties van de eigenschappen *frequentie*, *bekendheid* en *gebruik* was iets hoger dan die van *transparantie* en *visualiseerbaarheid*. We raden andere onderzoekers



aan om in het vervolg de D-coëfficiënt te gebruiken om de betrouwbaarheid van intuïties vast te stellen.

*Transparantie*, *bekendheid* en *visualiseerbaarheid* waren belangrijke voorspellers van idioomkennis. *Transparantie* en *bekendheid* hadden een positief effect op idioomkennis. Hoe bekender en transparanter een uitdrukking werd gevonden, hoe beter de proefpersonen in staat waren om de betekenis van de uitdrukking te selecteren in de meerkeuzevraag. *Visualiseerbaarheid* had een negatief effect op idioomkennis. Hoe gemakkelijker de proefpersonen een beeld konden vormen van de uitdrukking, hoe slechter hun idioomkennis was. Waarschijnlijk vormen de proefpersonen zich een beeld van de letterlijke interpretatie. Dit beeld kan de figuurlijke interpretatie in de weg zitten en kan daarom leiden tot meer fouten op de vraag naar de figuurlijke betekenis.

### **Hoofdstuk 3. Intuïties van idioomeigenschappen van T2-leerders**

Op dezelfde wijze als in Hoofdstuk 2 hebben we moedertaalsprekers van het Nederlands en Duitse T2-leerders van het Nederlands gevraagd om 110 Nederlandse idiomatische uitdrukkingen (een subset van de eerdere 374 uitdrukkingen) te beoordelen met betrekking tot dezelfde eigenschappen. Daarnaast hebben we ook weer de kennis van uitdrukkingen getoetst in een meerkeuzevraag en een open vraag. Aangezien T2-leerders in het algemeen weinig idiomatische uitdrukkingen kennen, vroegen we ons af of intuïties van T2-leerders wel een betrouwbare en bruikbare bron van informatie zijn. Daarnaast wilden we deze intuïties en de kennis van T2-leerders vergelijken met die van moedertaalsprekers. Richten T2-leerders zich meer op individuele woorden (Wray, 2002, 2008) of op de uitdrukking als geheel (Ellis, 2001)?

We vonden dat de intuïties over idioomeigenschappen van T2-leerders erg betrouwbaar waren, net als die van moedertaalsprekers. Daarnaast zagen we grote verschillen tussen T2-leerders en moedertaalsprekers in de oordelen op de eigenschappen die te maken hadden met de ervaring met de uitdrukkingen. Hoewel moedertaalsprekers hoog scoorden op de eigenschappen *frequentie*, *bekendheid* en *gebruik*, waren de scores van T2-leerders op deze

eigenschappen juist laag. De scores op *transparantie* en *visualiseerbaarheid* lagen veel dichterbij elkaar, maar moedertaalsprekers beoordeelden de 110 uitdrukkingen gemiddeld als minder transparant dan de T2-leerders, terwijl dit voor *visualiseerbaarheid* net andersom was.

*Transparantie* was de belangrijkste voorspeller van de idioomkennis van T2-leerders. Hoe transparanter de uitdrukking, hoe beter de idioomkennis. *Visualiseerbaarheid* had, net zoals bij de moedertaalsprekers in Hoofdstuk 2, een negatief effect op de idioomkennis van T2-leerders. Daarnaast bleken *gelijkenis met de moedertaal* en *woordenschat* de idioomkennis van T2-leerders positief te beïnvloeden.

De intuïties van T2-leerders bleken een belangrijke bron van informatie voor idioomonderzoek bij deze groep. Ook leek het erop dat T2-leerders, in vergelijking met moedertaalsprekers, meer gericht zijn op de individuele woorden dan op de uitdrukking als geheel. Dit betekent echter niet dat moedertaalsprekers en T2-leerders fundamenteel anders omgaan met idiomatische uitdrukkingen. Een mogelijke verklaring is dat T2-leerders, die veel minder ervaring hebben met de uitdrukkingen dan moedertaalsprekers, de uitdrukkingen als zodanig moeilijker kunnen herkennen. Om de figuratieve betekenis af te leiden, vertrouwen ze meer op de afzonderlijke woorden. Een mogelijke consequentie is dat T2-leerders die intensieve leerervaring opdoen met idiomen, meer op moedertaalsprekers kunnen gaan lijken in hun verwerking van idiomatische uitdrukkingen. Het leeraspect hebben we verder onderzocht in het volgende hoofdstuk.

#### **Hoofdstuk 4. Het leren van T2 idiomen in een CALL-omgeving**

Hoe kunnen verschillen in idioomkennis tussen T2-leerders en moedertaalsprekers worden verholpen door intensieve oefening met deze idiomen? Met andere woorden, gaat het hier inderdaad om een kwestie van ervaring? Dit is het onderwerp van Hoofdstuk 4. Ook hebben we hierin onderzocht of het helpt om tijdens oefeningen hardop te lezen in plaats van stil te lezen en hebben we de invloeden bestudeerd van *transparantie* en *gelijkenis met de moedertaal* op het leren van idiomatische uitdrukkingen.

Duitse leerders van het Nederlands oefenden met 60 Nederlandse idiomatische uitdrukkingen via een *Computer Assisted Language Learning* (CALL) systeem waarin ze automatisch voorzien werden van feedback. De helft van die 60 uitdrukkingen werd intensief geoefend en de andere helft maar in beperkte mate. Voor en na het oefenen werd de idioomkennis van de Duitse T2-leerders getoetst.

De intensiteit van oefenen had een duidelijk effect op het leren van de uitdrukkingen. T2-leerders scoorden in de nameting beter op uitdrukkingen die ze intensief geoefend hadden dan op uitdrukkingen die ze in beperkte mate hadden geoefend. De T2-leerders scoorden na intensief oefenen zelfs beter dan de moedertaalsprekers uit Hoofdstuk 2. Ook vonden we een duidelijk effect van *transparantie* en *gelijkenis met de moedertaal*. In de voormeting bleken T2-leerders hun moedertaal, het Duits, nog niet te gebruiken om de betekenis van de Nederlandse uitdrukkingen te bepalen. Na het oefenen deden ze dit echter wel. Uitdrukkingen die op elkaar leken en hetzelfde betekenden in de twee talen werden beter geleerd dan uitdrukkingen die niet op elkaar leken. Naarmate T2-leerders met de uitdrukkingen aan de slag gingen, werden ze zich wellicht bewust van de gelijkenissen. Dit zou verklaren waarom de leerders pas in de nameting effecten lieten zien van *gelijkenis met de moedertaal*. Hetzelfde zagen we gebeuren voor *transparantie*. In de voormeting was er geen effect van *transparantie* zichtbaar, maar in de nameting wel. T2-leerders hadden tijdens de voormeting nog niet in de gaten dat ze de individuele woorden konden gebruiken om de figuratieve betekenis af te leiden. Ze werden zich tijdens het oefenen bewust van deze waardevolle informatie en konden hier in de nameting wel gebruik van maken. Dit zagen we echter alleen voor uitdrukkingen die beperkt geoefend werden. We denken dat dit effect niet zichtbaar was bij uitdrukkingen die intensief geoefend werden omdat proefpersonen de betekenis van deze uitdrukkingen al kenden, omdat ze er zo veel mee geoefend hadden. Naast de effecten van intensiteit van oefenen, *transparantie* en *gelijkenis met de moedertaal*, vonden we geen effect van hardop lezen ten opzichte van stil lezen.

Dit hoofdstuk laat zien dat intensieve oefening met idiomen door middel van een CALL-systeem effectief is. T2-leerders scoorden na veel oefening even goed of zelfs beter dan moedertaalsprekers op een

kennistoets. Verschillen in idioomkennis tussen moedertaalsprekers en T2-leerders lijken dus inderdaad te wijten aan verschillen in ervaring. Met gerichte oefening kunnen deze verschillen worden weggenomen. Dit bevestigt ons idee dat moedertaalsprekers en T2-leerders dezelfde onderliggende mechanismen gebruiken om idiomen te leren. Hoeveel oefening nodig is, hangt af van de eigenschappen van de uitdrukkingen. Niet-transparante uitdrukkingen vergen meer oefening dan transparante uitdrukkingen. Dit geldt ook voor uitdrukkingen die niet lijken op uitdrukkingen in de moedertaal van de leerders.

### **Hoofdstuk 5. Idioomverwerking door moedertaalsprekers: de rol van de afzonderlijke woorden**

Neem de zin *De getrainde dief liep uiteindelijk toch tegen de lamp*. In zo'n context heb je de betekenissen van de losse woorden van de uitdrukking in principe niet nodig om de zin te begrijpen. Toch zul je, als je begint met het lezen van een zin, de betekenissen van de afzonderlijke woorden uit de uitdrukking proberen te begrijpen, omdat je nog niet weet dat het inderdaad om een uitdrukking gaat. Op het moment dat je de uitdrukking als zodanig herkent, heb je de individuele woordbetekenissen niet meer nodig. Je kunt de deze woordbetekenissen in principe negeren. In Hoofdstuk 5 onderzochten we of dit inderdaad zo werkt bij moedertaalsprekers.

Sterk sturende zinnen met uitdrukking, zoals hierboven, vergeleken we met zinnen zonder uitdrukking, maar met hetzelfde laatste woord. De tegenhanger van de idiomatische zin hierboven was bijvoorbeeld: *Het kind kan niet slapen zonder licht van een kleine lamp*. Beide zinnen eindigden met het woord *lamp*. In de eerste zin werd dit woord gebruikt als onderdeel van de uitdrukking, en in de tweede zin werd dit woord letterlijk gebruikt. Daarnaast werden deze bovenstaande zinnen ook af en toe gepresenteerd met een onverwacht ander laatste woord, bijvoorbeeld *De getrainde dief liep uiteindelijk toch tegen de kaars* of *Het kind kan niet slapen zonder licht van een kleine helm*. We vroegen proefpersonen om steeds het laatste woord van de zin, dat in het rood op het scherm verscheen, zo snel mogelijk hardop voor te lezen. We vergeleken hoe lang mensen erover deden om te reageren op de

verwachte en de niet-verwachte woorden indien ze figuurlijk of letterlijk gebruikt werden.

Op basis van hun reactietijden konden we vaststellen dat de betekenis van het laatste woord inderdaad genegeerd werd als het onderdeel was van een uitdrukking. Hoewel de betekenis van het laatste woord niet relevant was en onderdrukt werd, zagen we wel een effect van de vorm van het woord. Het woord *lamp* moest in de zin met de uitdrukking immers ook worden gelezen om te controleren of de zin inderdaad een uitdrukking bevatte.

Deze resultaten maken aannemelijk dat tijdens het verwerken van de uitdrukkingen door moedertaalsprekers zowel de vorm van de individuele woorden als de uitdrukking als geheel tegelijkertijd actief zijn. De vorm van deze woorden is cruciaal, omdat je op basis van deze informatie kan bepalen of het om een uitdrukking gaat. De betekenissen van de afzonderlijke woorden heb je in zo'n situatie echter niet nodig en zitten alleen maar in de weg als je de zin figuurlijk wil interpreteren. Dit pleit voor een hybride model van idioomverwerking. In zo'n model zijn de figuurlijke en de letterlijke betekenissen tegelijkertijd actief (Sprenger et al., 2006; Libben & Titone, 2008). Mede afhankelijk van de eigenschappen van de uitdrukking probeer je de losse woorden al dan niet letterlijk te interpreteren. Bij niet-transparante uitdrukkingen is de kans groter dat je de individuele woordbetekenissen negeert, zoals ook blijkt uit dit hoofdstuk. In hoeverre dit ook geldt voor T2-leerders hebben we onderzocht in Hoofdstuk 6.

### **Hoofdstuk 6. Idioomverwerking door T2-leerders: de rol van de afzonderlijke woorden en intensief oefenen**

Eerdere hoofdstukken lieten zien dat T2-leerders hun aandacht meer richten op de individuele woorden dan op de uitdrukking als geheel. De vraag is dan ook of dit tot een ander verwerkingspatroon leidt bij T2-leerders dan bij moedertaalsprekers. En wat is het effect van intensief oefenen op dit verwerkingsproces? Gaat het verwerkingsproces van T2-leerders door intensieve oefening meer lijken op dat van moedertaalsprekers, net zoals voor idioomkennis? Aangenomen dat T2-leerders zich in eerste instantie meer richten op de individuele woorden, kunnen ze dan na intensief oefenen met uitdrukkingen de individuele

woordbetekenissen negeren, net als moedertaalsprekers? Die individuele woordbetekenissen zijn immers niet nodig om niet-transparante uitdrukkingen te begrijpen.

Om deze vragen te beantwoorden hebben we het onderzoek van Hoofdstuk 5 nog eens uitgevoerd met de Duitse T2-leerders die eerder met het CALL-systeem hadden geoefend. De helft van de uitdrukkingen in het experiment werden intensief geoefend door de Duitse T2-leerders en met de andere helft van de uitdrukkingen oefenden de proefpersonen enkel in beperkte mate.

Zoals in Hoofdstuk 4 al aangetoond was, trad er een duidelijk oefeneffect op bij de idioomkennis van T2-leerders. Voorafgaand aan het oefenen konden de T2-leerders minder vaak de correcte betekenis herkennen dan na het oefenen. Voor de uitdrukkingen die intensief geoefend werden, was het verschil tussen de voor- en nameting het grootst. T2-leerders scoorden op die uitdrukkingen zelfs beter dan de moedertaalsprekers uit Hoofdstuk 2. Maar is een vergelijkbaar effect ook zichtbaar in de verwerking van deze idiomatische uitdrukkingen?

We vonden dat de afzonderlijke woordvormen uit de uitdrukkingen inderdaad actief werden, net zoals bij moedertaalsprekers. De woordvormen moesten immers herkend worden om na te gaan of het inderdaad om een uitdrukking ging. Dit proces werd echter niet beïnvloed door veelvuldig oefenen. Dat T2-leerders de uitdrukkingen wel als zodanig herkenden, bleek uit hun snellere reactie op het verwachte woord *lamp* in de zin *De getrainde dief liep uiteindelijk toch tegen de lamp*, dan op het onverwachte woord *helm* in de aangepaste tegenhanger van die zin (*De getrainde dief liep uiteindelijk toch tegen de helm*). Hier leek een effect van oefenen zichtbaar. Het verschil in reactietijd werd wat groter als T2-leerders van tevoren intensief geoefend hadden met de uitdrukking, maar het effect was echter niet statistisch significant. Daarnaast bleek het experiment niet gevoelig genoeg om vast te stellen of, net als bij moedertaalsprekers, de betekenissen van de afzonderlijke woorden werden genegeerd tijdens het lezen van uitdrukkingen.

Deze studie laat zien dat verschillen tussen T2-leerders en moedertaalsprekers qua idioomkennis konden worden weggenomen door intensief te oefenen met idiomatische uitdrukkingen. Hoewel dit

voor idioomverwerking minder duidelijk was, werd ook hier een klein effect van oefenen gevonden. Dit suggereert dat T2-leerders en moedertaalsprekers waarschijnlijk gebruik maken van dezelfde mechanismen om uitdrukkingen te verwerven.

Op basis van de resultaten in dit proefschrift stelden we een hybride idioomverwerkingsmodel voor, gebaseerd op eerder werk van Sprenger et al. (2006), dat voor moedertaalsprekers en T2-leerders hetzelfde werkt. Dit model verklaart verschillen en overeenkomsten tussen beide groepen aan de hand van verschillen in ervaring met idiomatische uitdrukkingen, de specifieke idioomeigenschappen en aspecten van de individuele woorden.

### **Conclusies**

In dit proefschrift hebben we idiomatische uitdrukkingen vanuit verschillende invalshoeken bestudeerd. Om te beginnen hebben we een grootschalige idioomdatabase opgezet, de eerste voor het Nederlands, en hebben we een maat geïntroduceerd om de betrouwbaarheid van intuïties te meten die nieuw was voor dit veld. Een belangrijke bron van informatie over idiomen vormden de intuïties van T2-leerders. We raden onderzoekers aan om meer aandacht te schenken aan deze intuïties. Vervolgens hebben we laten zien dat intensieve training met een CALL-systeem het leren van uitdrukkingen echt kan helpen. Vooral niet-transparante uitdrukkingen en uitdrukkingen die niet lijken op die in de moedertaal verdienen meer aandacht in het onderwijs. Nadere bestudering van de rol van ervaring, specifieke idioomeigenschappen en aspecten van de individuele woorden, leverden bewijs op voor een hybride model van idioomverwerking bij zowel moedertaalsprekers als T2-leerders. In zo'n model zijn bij de verwerking van idiomen de representaties voor idioombetekeningen (opgebouwd uit combinaties van woorden) in competitie met die van de letterlijke betekenis van de zin. Ondanks verschillen tussen de twee participantengroepen qua idioomkennis en verwerking, lijken de onderliggende mechanismen voor de verwerking van idiomen dezelfde te zijn. Aangetroffen verschillen lijken voornamelijk een gevolg van verschil in ervaring.

Ondanks zulke verschillen in ervaring met idiomemen, zijn idioomeigenschappen zoals *transparantie* en *visualiseerbaarheid* op dezelfde manier van invloed op de idioomkennis van T2-leerders en moedertaalsprekers. Wat de *verwerking* van idiomatische uitdrukkingen betreft werden zowel overeenkomsten als verschillen tussen T2-leerders en moedertaalsprekers zichtbaar. Zulke verschillen konden niet volledig weggenomen worden door intensieve oefening, maar, gezien de trend in ons onderzoek, zullen die verschillen naar verwachting geleidelijk aan verdwijnen bij nog intensievere idioom oefening door de T2-leerders.

Al met al heeft dit proefschrift aan het licht gebracht dat moedertaalsprekers en T2-leerders misschien dan wel als dag en nacht van elkaar verschillen in hun ervaring met idiomatische uitdrukkingen, maar één pot nat zijn als het gaat om de onderliggende mechanismen voor het onder de knie krijgen, begrijpen en gebruiken van idiomatische uitdrukkingen.



## **Curriculum Vitae**

Ferdy Hubers was born in Berlicum in 1991. He studied Linguistics at Radboud University and obtained his Bachelor's degree in 2012. He continued his studies at the same university and finished the research master in Language and Communication at the start of 2015. During his studies, Ferdy worked as a student assistant on various projects. In March 2015, Ferdy started working as a PhD candidate at the Centre for Language and Speech Technology (CLST), and the Centre for Language Studies (CLS) at Radboud University on a research project within the NWO funded project Idiomatic Second Language Acquisition (ISLA). During his PhD, he was asked to teach several courses within the Linguistics department, and co-supervised a number of Bachelor's and Master's students with theses and internships. He is currently working as a teacher at the department Language and Communication and as a post-doc at the CLST and CLS in the project Dutch Automatic Reading Tutor funded by NRO.



## List of publications

Hubers, F., Cucchiarini, C., & Strik, H. (submitted). Second language learner intuitions of idiom properties: Reliable indicators of idiom knowledge? *Studies in Second Language Acquisition*.

Cucchiarini, C., Hubers, F., & Strik, H. (under review). Learning L2 idioms in a CALL environment: the role of practice intensity, practice modality, and idiom properties. *Computer Assisted Language Learning*.

Hubers, F., Redl, T., de Vos, H., Reinartz, L. & de Hoop, H. (under review). Processing prescriptively incorrect comparative particles: evidence from sentence-matching and eye-tracking. *Frontiers in Psychology*.

Hubers, F., Cucchiarini, C., Strik, H., & Dijkstra, A.F.J. (2019). Normative Data of Dutch Idiomatic Expressions: Subjective Judgments You Can Bank on. *Frontiers in Psychology* 10:1075. doi: 10.3389/fpsyg.2019.01075

Hubers, F., Trompenaars, T., Collin, S., de Schepper, K. & de Hoop, H. (2019). Hypercorrection as a by-product of education. *Applied Linguistics*. doi: 10.1093/applin/amz001

Hubers, F., van Ginkel, W., Cucchiarini, C., Strik, H. & Dijkstra, A.F.J. (2018). Normative data on Dutch idiomatic expressions: Native speakers. DANS EASY [Dataset]. doi:10.17026/dans-zjx-hnsk.

Hubers, F., Snijders, T.M. & de Hoop, H. (2016). How the brain processes violations of the grammatical norm: An fMRI study. *Brain & Language* 163, pp. 22-31. doi: 10.1016/j.bandl.2016.08.006

Hubers, F. & de Hoop, H. (2013). The effect of prescriptivism on comparative markers in spoken Dutch. *Linguistics in the Netherlands 2013*, pp. 89-101.



## List of idioms on cover

The two idioms on the foreground:

Take the bull by the horns / de koe bij de hoorns vatten

Order from the upper left corner to the bottom right corner:

- |  |  |
|--|--|
| 1. Op de kleintjes letten  | 20. Met de gebakken peren<br>zitten                |
| 2. Van A tot Z   | 21. De bui zien hangen                             |
| 3. In de soep lopen  | 22. Van de hak op de tak<br>springen               |
| 4. Het regent pijpenstelen   | 23. Liefde maakt blind                             |
| 5. Zijn kop in het zand steken /<br>to bury your head in the<br>sand | 24. Het zwarte schaap                              |
| 6. Het zal hem een worst<br>wezen                                    | 25. Iemand een oor aannaaien                       |
| 7. Op de fles gaan   | 26. In de nesten zitten                            |
| 8. Dat muisje krijgt nog een<br>staartje                             | 27. Zich groen en geel ergeren                     |
| 9. De hond in de pot vinden  | 28. To kill two birds with one<br>stone            |
| 10. Op rozen zitten  | 29. Pigs might fly                                 |
| 11. Alle gekheid op een stokje                                       | 30. Don't put all your eggs in<br>one basket       |
| 12. Buiten zijn boekje gaan  | 31. Catch someone red-handed                       |
| 13. Iemand een poot uitdraaien<br>/ to pull someone's leg            | 32. A storm in a teacup                            |
| 14. Iemand een hart onder de<br>riem steken                          | 33. Shoot yourself in the foot                     |
| 15. Zijn eigen boontjes doppen                                       | 34. Bite the bullet                                |
| 16. Op hete kolen zitten   | 35. Kick the bucket                                |
| 17. Naast zijn schoenen lopen  | 36. Let the cat out of the bag                     |
| 18. De toon aangeven   | 37. Have a finger in the pie                       |
| 19. Het varkentje wassen   | 38. A bird in the hand is worth<br>two in the bush |
|  | 39. Paint the town red                             |

Illustrations by Kasper Boon (English Kaboons and Dutch Kaboontjes from 'Het leven is een feest, maar je moet zelf de slingers ophangen', Edition Albert Sickler; text Ingrid Regout and illustrations Kasper Boon).