

Dedicated Language Resources for Interdisciplinary Research on Multiword Expressions: Best Thing since Sliced Bread

Ferdy Hubers, Catia Cucchiari, Helmer Strik

Centre for Language Studies, Centre for Language and Speech Technology, Radboud University
Erasmusplein 1, 6525 HT, Nijmegen, the Netherlands
{f.hubers,c.cucchiari,w.strik}@let.ru.nl

Abstract

Multiword expressions such as idioms (*beat about the bush*), collocations (*plastic surgery*) and lexical bundles (*in the middle of*) are challenging for disciplines like Natural Language Processing (NLP), psycholinguistics and second language acquisition, due to their more or less fixed character. Idiomatic expressions are especially problematic, because they convey a figurative meaning that cannot always be inferred from the literal meanings of the component words. Researchers acknowledge that important properties that characterize idioms such as frequency of exposure, familiarity, transparency, and imageability, should be taken into account in research, but these are typically properties that rely on subjective judgments. This is probably one of the reasons why many studies that investigated idiomatic expressions collected limited information about idiom properties for very small numbers of idioms only. In this paper we report on cross-boundary work aimed at developing a set of tools and language resources that are considered crucial for this kind of multifaceted research. We discuss the results of our research and suggest possible avenues for future research.

Keywords: Multiword expressions, language resources, interdisciplinary research, idioms

1. Introduction

Multiword expressions (MWEs), also known as formulaic expressions, are recurrent sequences of words that tend to co-occur in more or less fixed combinations. These MWEs come in different types, such as idiomatic expressions (*hit the sack*), collocations (*make a decision*, *plastic surgery*), binomials (*back and forth*), lexical bundles (*in the middle of*), and speech formulas (*I beg your pardon*) (Cacciari, 2014; Nattinger & DeCarrico, 1992; Siyanova-Chanturia & Martinez, 2014).

MWEs in general, and idiomatic expressions in particular, are challenging from the perspective of Natural Language Processing (NLP), psycholinguistics and second language acquisition (SLA), and are therefore widely studied in these fields.

From the perspective of NLP, MWEs are challenging because they pose serious difficulties in terms of processing in language and speech technology. Systems that use only compositional methods to generate text without taking into account the special characteristics of MWEs, can result in the generation of non-native-like constructions such as *plastic operation*. This is called the *overgeneration problem* (Sag et al., 2002). Expressions that convey a figurative meaning that cannot be extracted from the individual word meanings, can lead to parsing problems. These difficulties are often referred to as the *idiomaticity problem* (Sag et al., 2002). Both problems are challenging for applications that rely on semantic information, such as systems for automatic translation and automatic summarization. In order to overcome the overgeneration and idiomaticity problem, many researchers have been trying to automatically identify MWEs and more specifically idiomatic expressions (e.g., Constant et al., 2017; Cruys & Villada Moirón, 2007; Odijk, 2013; Ramisch, 2015; Savary et al., 2015; Strik, Hulsbosch, & Cucchiari, 2010; Wahl & Gries, 2018).

From the perspective of psycholinguistics and SLA, MWEs are challenging because of their (semi-) fixed character. For

this reason, second language (L2) learners have enormous difficulties to acquire these expressions (Cieślicka, 2006; Conklin & Schmitt, 2008; Ellis, Simpson-Vlach, & Maynard, 2008; Wray, 2000). Idiomatic expressions are especially problematic because they convey a figurative meaning that often cannot be derived from the literal meanings of the component words (Cacciari, 2014). As a consequence, psycholinguistics investigates to what extent these expressions are processed just like literal strings of words, or whether they are stored as a whole in the mental lexicon (Bobrow & Bell, 1973; Cacciari & Tabossi, 1988; Carrol & Conklin, 2014; Cieślicka, 2006, 2013; Sprenger, Levelt, & Kempen, 2006). In SLA, researchers have mainly focused on the difficulties L2 learners experience with the comprehension and production of L2 idioms and other types of MWEs (Irujo, 1986, 1993; Kellerman, 1979; Pawley & Syder, 1983), and on the effectiveness of new teaching and learning methods to improve L2 idiom comprehension and production (Boers et al., 2006; Boers, Eyckmans, & Stengers, 2007; Boers & Lindstromberg, 2012; Zyzik, 2011).

Researchers from the fields described above acknowledge that it is important to take into account a number of relevant properties that characterize idioms and that determine their relative difficulty in processing and acquisition. Idiom properties, such as *frequency of exposure*, *familiarity*, *transparency*, and *imageability*, have all been found to affect idiom processing and acquisition to different degrees. In NLP, MWEs are often divided into two groups depending on their transparency or compositionality, but these are typically properties that rely on subjective judgments. This is probably one of the reasons why many studies that investigated idiomatic expressions, only collected limited information about idiom properties for very small numbers of idioms only. However, this is problematic for research, because it hinders comparability of results. Moreover, conclusions can be drawn only for specific sets of idioms.

This all seems to suggest that different fields of investigation like NLP, psycholinguistics and applied

linguistics, could profit from large databases of MWEs in general and idioms in particular, in which information about the most important properties has been gathered through systematically obtained subjective judgments and more objective measurements. At present, only a limited number of such resources have been compiled for idioms (Bonin, Méot, & Bugajska, 2013; Citron et al., 2016; Nordmann & Jambazova, 2017; Tabossi, Arduino, & Fanari, 2011). Moreover, the databases available often differ on the definitions of the idiom properties that are assessed, their nature and number, the number of idioms included in the database, and the way the data is collected.

In this paper, we report on cross-boundary work we conducted within the framework of an interdisciplinary project on formulaic language aimed at studying the acquisition and processing of MWEs from different perspectives in psycholinguistics and second language acquisition, the ISLA project¹. In this project we developed a set of tools and language resources that were considered crucial for this kind of multifaceted research. These resources include the following:

a) A large database of 374 Dutch idiomatic expressions for which we collected essential information on the properties of these idioms, all obtained from more than 350 native speakers (Hubers, Cucchiari, Strik, & Dijkstra, 2019; Hubers, van Ginkel, Cucchiari, Strik, & Dijkstra, 2018). For a subset of 110 idiomatic expressions, we collected data from German L2 learners of Dutch, and for a subset of 60 idiomatic expressions data was obtained from Arabic L2 learners of Dutch (Cucchiari, Hubers, & Strik, 2019; Hubers, Cucchiari, & Strik, 2019).

b) A test of idiom knowledge and idiom properties that was used to assess the idiom knowledge of native speakers and L2 learners and to collect information about idiom properties. Through this test, we gathered data to be included in the idiom database (Cucchiari et al., 2019; Hubers, Cucchiari, & Strik, 2019).

c) A Computer Assisted Language Learning (CALL) system with immediate corrective feedback that was used to provide L2 learners with the opportunity to practice Dutch idiomatic expressions (Cucchiari, Hubers, & Strik, 2020).

The paper is organized as follows. In Section 2, we describe the resources that we developed within our interdisciplinary research project on MWEs. Section 3 provides an overview of the results related to the various resources. In Section 4 we discuss the importance of such resources for interdisciplinary research and provide suggestions for future developments.

2. Resources

2.1 Idiom Database

Many studies on idiomatic expressions are based on small numbers of idioms with only limited information about their properties (e.g., Cacciari & Tabossi, 1988; Cutting & Bock, 1997; Holsinger, 2013; Siyanova-Chanturia, Conklin, & Schmitt, 2011), but this has clearly

shortcomings for research, because of the multifaceted nature of idiomatic expressions.

For this reason we compiled a large-scale database of Dutch idiomatic expressions, which served as a firm basis for our research on idiomatic expressions. The expressions included in this database were collected from different sources such as DuELME (Odijk, 2013), Stoett, (1925), and websites like woorden.org (Slot Webcommerce BV, 2017) and OnzeTaal (Genootschap OnzeTaal, 2011). The expressions selected are idioms and do not include other types of formulaic expressions like collocations, binomials and lexical bundles (Siyanova-Chanturia & Martinez, 2014). For 374 idioms, we collected information about perceived frequency of exposure, meaning familiarity, usage, transparency and imageability from 394 native speakers. In addition to these subjective measures, we gathered objective data on corpus frequency from the SoNaR corpus of written Dutch (Oostdijk et al., 2013), consisting of 500 million words. For a subset of 110 and 60 idiomatic expressions we collected data on idiom properties from, respectively, German and Arabic L2 learners of Dutch. Moreover, we asked bilinguals to assess the extent to which some Dutch idioms exist in German and Standard Arabic.

This database has been a crucial element in our research on idiom processing and acquisition. It provided important insights into the knowledge, use and subjective judgments of idiom properties of Dutch idioms by native speakers and L2 learners. Moreover, based on this database, we were able to make informed decisions about which idioms to include in our psycholinguistic experiments and learning experiments. 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 part of the database with information from native speakers has already been made available (Hubers et al., 2018). The additional data on language learners and Dutch emigrants (see below) will be disseminated very soon.

2.2 Test of Idiom Knowledge and Idiom Properties

This test is a web-based test that has been used to collect data on idiom knowledge and subjective judgments of idiom properties from native speakers and L2 learners. To assess idiom knowledge two procedures were employed:

- a) a multiple choice test of meaning recognition developed for all idioms from the database, and
- b) an open question of meaning recall for the same idioms from the database.

Participants were presented with four options in the multiple choice test. We created three incorrect alternative meanings that would be plausible if one were not familiar with the idiom. Subjective judgments about the idiom properties *frequency of exposure*, *meaning familiarity*, *usage*, *transparency*, and *imageability* were collected through five-point Likert scales. We used the following (working) definitions:

¹ <http://isla.ruhosting.nl>

- Frequency of exposure - the relative degree to which participants indicate they have come across an idiomatic expression in speech or in print (Titone & Connine, 1994).
- Usage - the frequency with which participants indicate they have used an idiomatic expression.
- Familiarity - how well speakers say they know the meaning of an idiom (Nordmann, Cleland, & Bull, 2014).
- Imageability - the extent to which an idiom can evoke an image (in line with Steinel, Hulstijn, & Steinel, 2007).
- Transparency - 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 included a background questionnaire in order to collect metadata. The version of the test for L2 learners contained more detailed language background questions about their exposure to and experience with Dutch than the version for the native speakers. Moreover, the test for L2 learners also included the LexTale vocabulary test (Lemhöfer & Broersma, 2012), as a more objective indicator of the L2 learner's proficiency in Dutch than self-reported data (van Ginkel, Hubers, Cucchiari, Dijkstra, & Strik, 2016).

Information on idiom knowledge is crucial in research on idiom processing in both native speakers and L2 learners. If participants do not know the idioms under study, possible differences in idiom processing are difficult to interpret. Using a test of idiom knowledge 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 (Beck & Weber, 2016; Cieśllicka, 2013; Nordmann et al., 2014; Titone & Connine, 1994). As a matter of fact, this method which was proposed in Hubers, van Ginkel, Cucchiari, Dijkstra, & Strik (2016) and Van Ginkel et al. (2016) for Dutch idioms was later adopted by Carrol, Littlemore, & Gillon Dowens (2017) in their study on native and non-native understanding of figurative phrases from English, German, Bulgarian and Chinese.

This idiom knowledge test was administered to assess the receptive knowledge of Dutch idiomatic expressions by participants in pre-tests and post-tests as part of our learning experiments (Cucchiari et al., 2020). In addition, this test was later used in an adapted form to test the idiom knowledge of Arabic L2 learners (Cucchiari et al., 2019; Hubers, Cucchiari, & Strik, 2019) and that of Dutch and Flemish emigrants to various countries in the framework of a study on the preservation of the Dutch language and culture by different groups of emigrants (see Doreleijers, van der Sijs, Assendelft, & Ooijevaar, 2019).

2.3 CALL System

We developed a CALL system to be able to systematically investigate idiom learning under various conditions of practice. For this CALL system, we designed training sessions intended to practice Dutch idiomatic expressions through a variety of paradigms and exercises (Cucchiari et al., 2020).

The first training session constituted a Paired Associate Learning (PAL) paradigm (Steinel et al., 2007), and a gap-fill exercise (see Figures 1 and 2 respectively). In the PAL exercise, each idiom was presented at the center of the screen with the corresponding meaning directly below. After 30 seconds the next idiom-meaning pair would appear on the screen. Participants were instructed to read the idioms and their meanings. In the gap-fill exercise idioms were embedded in a sentence in which one word was missing. Participants were asked to fill in the missing word. This word was always part of the idiom.



Figure 1: A screenshot of the PAL exercise



Figure 2: A screenshot of the gap-fill exercise

The second and third training sessions consisted of a sentence completion task (Figure 3) and an idiom selection exercise (Figure 4). In the sentence completion task, given sentences had to be completed with the appropriate idiomatic expression (multiple choice). In the idiom selection exercise, participants had to select the idiomatic expression that matched the given figurative meaning (multiple choice).



Figure 3: A screenshot of the sentence completion exercise



Figure 4: A screenshot of the idiom selection exercise

In the gap-fill, sentence completion, and idiom selection exercises, L2 learners received automatic corrective feedback immediately after they had provided 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.

An important aspect of this CALL system is that all system-learner interactions were logged. This information allows us to get more insight into learner behavior. All utterances by the users are recorded and are thus available for subsequent research. We stored interaction data that can be relevant for research purposes, for instance for studying the effects of practice and corrective feedback on performance and proficiency.

Using this kind of system for research on idiom learning has several important advantages. First, it allows researchers to control for and systematically vary a number of crucial variables such as language materials, type and intensity of practice and type, amount and timing of feedback, which would not be possible in classroom situations. This does not mean, however, that the CALL environment becomes unrealistic. On the contrary, a second important advantage is that a CALL system environment has high ecological validity. The 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 of intensive and individualized practice that would not otherwise be feasible in classroom instruction.

3. Results

3.1 Idiom Database

Previous research had produced mixed results with respect to the reliability of subjective judgments of idiom properties, both by native speakers and L2 learners (Bonin et al., 2013; Citron et al., 2016; Tabossi et al., 2011; Titone & Connine, 1994). Since reliability is an essential property of measurements that has to be assessed before using the data for further research, we calculated several metrics of reliability for our data. We found that the subjective judgments of native speakers about the idiom properties of the 374 Dutch idioms in our database were highly reliable as measured by the D-coefficient (D-coefficient > 0.87) (Hubers et al., 2019). This reliability measure was chosen instead of other more common measures like Cronbach's

alpha and the Intra-Class Correlation Coefficient (ICC), because it has important advantages, as it can handle sources of variance other than items and raters, different research designs like unbalanced research designs, and because it allows to calculate the minimum number of raters required to obtain reliable data (Li et al., 2015; Shavelson & Webb, 2006). The reliability of the subjective judgments by L2 learners for a subset of 110 Dutch idiomatic expressions from the database was assessed using the ICC as in this case we had a fully crossed design. The German L2 judgements also appeared to be highly reliable (ICC > 0.86). For the subset of 60 idioms for which we obtained subjective judgments from both L1 German and Arabic L2 learners of Dutch, the results indicate slightly lower reliability as measured by the ICC. On the one hand this seems obvious because the test contained fewer items (almost half). On the other, we see that while for the L1 German learners reliability is still high, except for usage, for the L1 Arabic learners reliability is much lower (ICC between 0.40 and 0.64). This means that in principle both native speakers and L2 learners are capable of assessing these idiom properties in a reliable way, but that the results can be dependent on important factors such as language distance and L2 proficiency. It is important to check the reliability before judgments are used as a basis for further research on idiom processing and acquisition.

Since the idiom database contains information about which idiomatic expressions are considered to be familiar and opaque to native speakers, these data could be used to select idiomatic expressions for our psycholinguistic experiments and learning experiments. For example, we only included idioms in our learning experiments that are well known by native speakers of Dutch. Table 1 presents five idiomatic expressions that were rated as highly familiar by native speakers, and therefore were included (among 55 other idioms) in our learning experiment. As can be seen from the same table, the familiarity intuitions of German and Arabic L2 learners for most of these idioms are much lower.

Idiomatic expression	L1	German L2	Arabic L2
voor aap staan	4.52	1.58	1.28
voor spek en bonen	4.42	1.40	1.18
iets uit je duim zuigen	4.35	3.37	1.43
iets onder de knie krijgen	4.21	2.55	3.24
voor iemand door het vuur gaan	4.19	3.85	1.27

Table 1: Mean familiarity scores (scale 1-5) for five Dutch idiomatic expressions by native speakers (L1) and L2 learners.

In a psycholinguistic study on idiom processing we wanted to include only idioms that are considered opaque to investigate the role of the individual words during the processing of these types of idioms. Examples of relatively familiar and opaque idiomatic expressions that we included in our psycholinguistic experiments are *veel voeten in de aarde hebben* (lit. 'to have many feet in the earth', meaning

‘to have difficulties with something’) and *koek en ei zijn* (lit. ‘to be a cookie and egg’, meaning ‘to be good friends’). Table 2 shows five highly opaque idiomatic expressions as rated by native speakers. The average scores of the German and Arabic L2 learners are also provided as a comparison.

Idiomatic expression	L1	German L2	Arabic L2
iets soldaat maken	1.31	1.74	1.15
boter bij de vis	1.31	1.37	1.23
het op zijn heupen krijgen	1.48	1.58	1.34
een wassen neus	1.57	2.00	1.51
iemand iets in de maag splitsen	1.83	2.53	1.30

Table 2: Mean transparency scores (scale 1-5) for five opaque Dutch idiomatic expressions by native speakers (L1) and L2 learners.

3.2 Test of Idiom Knowledge and Idiom Properties

The test of idiom knowledge and idiom properties has been administered to different types of participants. Table 3 presents an overview of the characteristics of the participant groups. The majority of participants were Dutch native speakers living in the Netherlands, but we also tested German and Arabic L2 learners of Dutch. In addition, for a study on Dutch language use by Dutch emigrants, an adapted version of the test was developed, which was filled in by 796 participants. These data are not reported here, but will become available soon.

	L1	German L2	Arabic L2
Age	20.40 (1.50)	24.76 (3.46)	24.50 (5.75)
LexTale (1-100)	NA	69.04 (11.75)	58.60 (5.90)
N	394	44	30

Table 3: Overview of the age and LexTale scores for native speakers of Dutch (L1), German L2 learners of Dutch, and Arabic L2 learners of Dutch. SDs are presented in brackets.

For the L2 speakers, we also collected more detailed information about their language background, as presented in Table 4.

	German L2	Arabic L2
Self-reported proficiency (scale 1-7)	5.51 (1.11)	4.68 (0.71)
#hours a week speaking Dutch outside class or work	9.66 (17.46)	8.54 (9.86)
#years learning Dutch	3.65 (2.38)	1.66 (1.02)
#years living in the Netherlands	1.45 (2.29)	2.44 (0.84)

Table 4: Detailed overview of language background variables for German and Arabic L2 learners of Dutch. SDs are presented in brackets.

The idiom knowledge of the three participant groups as assessed by the multiple choice test on idiom meaning recognition are shown in Table 5. The idiom knowledge is presented for the same set of idiomatic expressions ($n = 60$).

	Idiom knowledge
L1 speakers	0.77 (0.24)
German L2 learners	0.55 (0.50)
Arabic L2 learners	0.36 (0.48)

Table 5: Mean proportion correct (and SDs) on the multiple choice test on idiom meaning recognition for 60 idioms.

3.3 CALL System

The CALL system was developed and used to study idiom learning. Since idioms are generally considered to be difficult to learn, we wanted to investigate whether focused and intensive practice as can be provided through a CALL system is effective for acquiring L2 idioms. Our research (Cucchiari et al., 2020) shows that L2 learners can indeed learn L2 idioms if they receive sufficient practice, something that is not always possible in L2 classrooms. The amount of practice required turns out to differ depending on the specific properties of idioms and the learners’ language background. In turn, knowledge about the idiom properties can therefore be used to design effective teaching interventions for specific L2 learners. In a CALL system all these variables can be systematically manipulated so as to achieve maximum flexibility and learning outcome. For instance, we found that practice with the CALL system facilitated idiom knowledge for both German L2 learners of Dutch (Cucchiari et al., 2020), and Arabic L2 learners of Dutch (Cucchiari et al., 2019). Although the focus in Cucchiari et al. (2020) was on investigating the effect of intensity and modality of practice and idiom properties on L2 idiom learning, these results allow to compare the idiom knowledge of the German L2 learners prior to CALL-based training (pre-test) with their idiom knowledge of the same idioms after training (post-test). In Table 6 we present these data together with those obtained for L2 learners with Arabic as their native language in the pre-test and post-test.

L2 learners	Test	Intensity of Practice	
		Limited	Intensive
German	Pre-test	0.55 (0.50)	0.56 (0.50)
	Post-test	0.65 (0.48)	0.83 (0.38)
Arabic	Pre-test	0.37 (0.48)	0.36 (0.48)
	Post-test	0.47 (0.50)	0.62 (0.49)

Table 6: Mean proportion correct (and SDs) on the Dutch idiom knowledge test at pre- and post-test by intensity of practice for German and Arabic L2 learners of Dutch

4. Discussion

The results presented above indicate that dedicated language resources can enhance research on idiom processing and acquisition and facilitate cross-fertilization among related disciplines. The cross-boundary character of this work appears from the fact that this initiative combined knowledge, methods and procedures from

psycholinguistics, SLA, CALL and computational linguistics. The language resources developed so far have partly already been made available for further research, and this process will continue in the future, as we get more data on L2 learners. In addition, the idiom database can be enriched with information on processing deriving from psycholinguistic experiments and data on acquisition deriving from the learning studies.

At the international workshop FLIPR² that was organized in Nijmegen in 2018 within the framework of the interdisciplinary NWO project ISLA, these results were presented and discussed with an interdisciplinary group of experts. These experts were formulaic language researchers from psycholinguistics, SLA and computational linguistics, some of whom presented related research for other languages. Data on formulaic language are difficult to obtain, and due to their generally limited size, multilingual character and dedicated purpose, they should be combined to be suitable for interdisciplinary research. Moreover, they are also very costly to collect. During the discussion and the brainstorming session that was held at the end of the workshop the need emerged to bring together existing and new data on MWEs in multiple languages in an interoperable and consistent way that would serve the requirements of the various research communities involved. A similar initiative was undertaken by the PARSEME Cost Action³ in the field of NLP. In SLA and psycholinguistics this has not been done before. Ideally such a cross-linguistic enterprise should be organized under the auspices of a larger research infrastructure initiative such as CLARIN ERIC. The CLARIN infrastructure is regarded as indispensable for this purpose. The data can be archived at local CLARIN centers whereas they can be made findable through a central portal via their (harvested) metadata. CLARIN precisely offers the standards, best practices and services which are needed for this.

For this reason contact was sought with developers of previous databases (Bonin et al., 2013; Citron et al., 2016; Nordmann & Jambazova, 2017; Tabossi et al., 2011) and other MWE researchers using methods from SLA, psycholinguistics, computational linguistics and CALL, which has led to a CLARIN ERIC project⁴. The aim of the project is to bring together an interdisciplinary group of experts in formulaic language (FL) research to set up a plan to collect existing and new data and to include these in the CLARIN infrastructure. Research will address methods for identifying MWEs that go beyond the simple frequency-based approaches, procedures for data collection that ensure uniformity and comparability of results and the development of a database of available data that can be continuously enriched and expanded with new data.

5. Conclusions

Developing dedicated language resources for interdisciplinary research on MWEs is necessary for conducting sound research and for allowing comparability of results. Making these resources available is important because data on MWEs in general, and idioms in particular are difficult to obtain and expensive to collect. CLARIN ERIC offers possibilities to start realizing a multilingual

database and research infrastructure to advance interdisciplinary research on MWEs.

6. Acknowledgements

We are grateful to our fellow members of the ISLA team, Ton Dijkstra and Wendy van Ginkel who contributed to the database and the test, and 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 was financed by the Dutch Research Council (NWO). We thank CLARIN ERIC for supporting infrastructural initiatives for formulaic language research.

7. Bibliographical References

- Beck, S. D., & Weber, A. (2016). 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.
- Bobrow, S. a., & Bell, S. M. (1973). On catching on to idiomatic expressions. *Memory & Cognition*, 1(3), 343–346.
- 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.
- 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.
- Boers, F., & Lindstromberg, S. (2012). Experimental and Intervention Studies on Formulaic Sequences in a Second Language. *Annual Review of Applied Linguistics*, 32, 83–110.
- Bonin, P., Méot, A., & Bugaiska, A. (2013). Norms and comprehension times for 305 French idiomatic expressions. *Behavior Research Methods*, 45, 1259–1271.
- Cacciari, C. (2014). Processing multiword idiomatic strings: Many words in one? *The Mental Lexicon*, 9(2), 267–293.
- Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, 27, 668–683.
- 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.
- Carrol, G., Littlemore, J., & Gillon Dowens, M. (2017). Of false friends and familiar foes: Comparing native and non-native understanding of figurative phrases. *Lingua*.
- Cieślicka, A. B. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research*, 22(2), 115–144.
- 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.

² <http://isla.ruhosting.nl/workshop>

³ <https://typo.uni-konstanz.de/parseme/index.php/the-action>

⁴ <http://hstrik.ruhosting.nl/FLIPR/>

- 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.
- 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.
- Constant, M., Eryiğit, G., Monti, J., van der Plas, L., Ramisch, C., Rosner, M., & Todirascu, A. (2017). Multiword Expression Processing: A Survey. *Computational Linguistics*, 43(4), 837–892.
- Cruys, T. Van de, & Villada Moirón, B. (2007). Semantics-based multiword expression extraction. *Proceedings of the Workshop on A Broader Perspective on Multiword Expressions*, (June), 25–32.
- Cucchiari, C., Hubers, F., & Strik, H. (2019). Learning idioms in a second language through a CALL system: the effect of transparency and language distance. In *Paper presented at the CALL Conference 2019*. Hong Kong.
- Cucchiari, C., Hubers, F., & Strik, H. (2020). Learning L2 idioms in a CALL environment: The role of practice intensity, modality, and idiom properties. *Computer Assisted Language Learning*.
- 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.
- Doreleijers, K., van der Sijs, N., Assendelft, B., & Ooijevaar, E. (2019). *Onderzoeksrapport Vertrokken Nederlands*. <https://taalunie.org/publicaties/92/onderzoeksrapport-vertrokken-nederlands>
- 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.
- Genootschap OnzeTaal. (2011). *Uitdrukkingen en Spreekwoorden*. <https://onzetaal.nl/taaladvies/thematisch/uitdrukkingen-en-spreekwoorden/>
- Holsinger, E. (2013). Representing Idioms: Syntactic and Contextual Effects on Idiom Processing. *Language and Speech*, 56(3).
- Hubers, F., Cucchiari, C., & Strik, H. (2019). Intuitions and knowledge of idiomatic expressions by L2 learners: The role of cross-language similarity. *Presentation at VOCAB*. Leuven, Belgium.
- 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.
- Hubers, F., van Ginkel, W., Cucchiari, C., Dijkstra, T., & Strik, H. (2016). How Native Speakers see the Light. *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].
- Irujo, S. (1986). 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.
- Kellerman, E. (1979). Transfer and Non-Transfer: Where We Are Now. *Studies in Second Language Acquisition*, 2(1), 37–57.
- Kovecses, Z., & Szabó, P. (1996). Idioms: A View from Cognitive Semantics. *Applied Linguistics*, 17(3), 326–255.
- 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.
- 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.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical Phrases and Language Teaching*.
- 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.
- 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>
- Odiijk, J. (2013). Identification and Lexical Representation of Multiword Expressions. In P. Spyns & J. Odiijk (Eds.), *Essential Speech and Language Technology for Dutch* (pp. 305–319). Springer. <http://doi.org/10.1007/978-3-642-30910-6>
- 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.
- 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.
- Ramisch, C. (2015). *Multiword Expressions Acquisition: A Generic and Open Framework* (Vol. XIV). Springer.
- Sag, I., Baldwin, T., Bond, F., Copestake, A., & Flickinger, D. (2002). Multiword Expressions: A Pain in the Neck for NLP. In *Proceedings of the 3rd International Conferences on Intelligent Text Processing and Computational Linguistics* (pp. 1–15).
- Savary, A., Sailer, M., Parmentier, Y., Rosner, M., Rosen, V., Przepirkowki, A., ... Sangati, F. (2015). PARSEME - PARSing and Multiword Expressions within a European multilingual network. *Proceeding of the 7th Language & Technology Conference (LYC 2015)*, 438–444.
- 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.
- 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.
- Siyanova-Chanturia, A., & Martinez, R. (2014). The Idiom Principle Revisited. *Applied Linguistics*, 1–22.
- Slot Webcommerce BV. (2017). Nederlands Woordenboek - Gezegden. <https://www.woorden.org>
- Sprenger, S., Levelt, W., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, 54(2), 161–184.
- 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.
- Stoett, F. A. (1925). *Nederlandse spreekwoorden, spreekwijzen, uitdrukkingen en gezegden* (4th edition). Zutphen, the Netherlands: W.J. Thieme & Cie.
- Strik, H., Hulsbosch, M., & Cucchiari, C. (2010). Analyzing and identifying multiword expressions in spoken language. *Language Resources and Evaluation*, 44, 41–58.
- Tabossi, P., Arduino, L., & Fanari, R. (2011). Descriptive norms for 245 Italian idiomatic expressions. *Behavior Research Methods*, 43(1), 110–123.
- 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.
- van Ginkel, W., Hubers, F., Cucchiari, C., Dijkstra, T., & Strik, H. (2016). Norming Studies for Idiom Processing: Native and Non-Native Benchmarks. *Paper presented at FLARN 2016*.
- Wahl, A., & Gries, S. T. (2018). Multi-word Expressions: A Novel Computational Approach to Their Bottom-Up Statistical Extraction (pp. 85–109). Springer, Cham.
- Wray, A. (2000). Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics*, 21(4), 463–489.
- Zyzik, E. (2011). Second language idiom learning: The effects of lexical knowledge and pedagogical sequencing. *Language Teaching Research*, 15(4), 413–433.