

Radboud University



WORKSHOP

Formulaic Language Processing and acquisition Research (FLIPR)

June 18-19, 2018

Radboud University Nijmegen

The Netherlands



ISLA

Idiomatic Second Language Acquisition

DONDERS
INSTITUTE



Centre for Language Studies
Radboud University



**Centre for Language and
Speech Technology**
Radboud University



NWO

Netherlands Organisation for Scientific Research

PROGRAM

Day 1 – Monday June 18th (all presentations 20' + 5')

08.45 – 09.15	Registration and Coffee & Tea
09.15 – 09.25	Welcome by Helmer Strik & Ton Dijkstra
09.30 – 10.45	Session 1: Resources
09.30 – 09.55	1. <i>Lut Colman</i>
09.55 – 10.20	2. <i>Alessandro Lenci</i>
10.20 – 10.45	3. <i>Ferdy Hubers</i>
10.45 – 11.05	Session Discussion
11.05 – 11.20	Coffee & Tea
11.20 – 12.10	Session 2: L1 Processing
11.20 – 11.45	4. <i>Cristina Cacciari</i>
11.45 – 12.10	5. <i>Robert Grimm</i>
12.10 – 12.30	Session Discussion
12.30 – 13:30	Lunch
13.30 – 14.20	Session 3.1: Acquisition 1
13.30 – 13.55	6. <i>Elke Peters</i>
13.55 – 14.20	7. <i>Catia Cucchiarini</i>
14.20 – 14.40	Coffee & Tea
14.40 – 15.30	Session 3.2: Acquisition 2
14.40 – 15.05	8. <i>Kathy Conklin</i>
15.05 – 15.30	9. <i>Simone Sprenger</i>
15.30 – 15.50	Session Discussion
16.05 – 16.30	General Discussion
16.30 – 17.30	Drinks
19.00	Workshop Dinner

Day 2 - Tuesday June 19th (all presentations 20' + 5')

- 08.45 – 09.00 **Coffee & Tea**
- 09.00 – 09.50 **Session 4.1: L2 Processing 1**
09.00 – 09.25 10. *Sara Beck & Andrea Weber*
09.25 – 09.50 11. *Wendy van Ginkel*
- 09.50– 10.05 **Coffee & Tea**
- 10.05 – 10.55 **Session 4.2: L2 Processing 2**
10.05 – 10.30 12. *Karolina Rataj*
10.30 – 10.55 13. *Ferdy Hubers*
- 10.55 – 11.15 **Session Discussion**
- 11.15 – 11.30 **Coffee & Tea**
- 11.30 – 12.20 **Session 5: Modelling**
11.30 – 11.55 14. *Rüdiger Thul*
11.55 – 12.20 15. *Louis ten Bosch*
- 12.20 – 12.40 **Session Discussion**
- 12.40 – 13:00 **General Conclusions & Wrap Up** by Ton Dijkstra & Helmer Strik
- 13.00 – 14.00 **Lunch & Farewell**

ORAL SESSIONS

1. Word Combinations

Speaker: *Lut Colman*

Institute for Dutch Lexicology, the Netherlands

Monday June 18th Session 1: 09.30-09.55

Phraseological information should be available to users, learners and teachers of Dutch and it should be rapidly and easily accessible. Online dictionaries are an obvious means to this end. However, on the one hand, phraseological information in general language dictionaries is scattered over layered microstructures and hence difficult to access. On the other hand, the very few Dutch phraseological dictionaries focus on particular phraseological (sub)types, for example collocations or idioms, while excluding other types. As a result of this fragmented phraseological landscape, learners of Dutch are dependent on many different resources to meet various phraseological needs, which is an impractical and undesirable learning environment.

The project *Woordcombinaties* (Word Combinations) of the Dutch Language Institute will be a new online lexicographic resource which can answer various types of phraseological queries by combining access to collocations, idioms and patterns in one tool.

In this contribution we present the three-year pilot during which we design the project and start with the description of the combinatorics of a selection of verbs for advanced learners of Dutch as a second language. We will merge a pattern dictionary of Dutch verbs, following the example of the *Pattern Dictionary of English Verbs (PDEV)*, with a collocation application, following the example of *Sketch Engine for Language Learning (SkeLL)*. This combination will make it possible for language learners to have quick access to collocations and to the pattern-meaning pairs in which these collocations occur.

The project is corpus based and the *Sketch Engine* will be used as corpus query system. *Patrick Hanks' Corpus Pattern Analysis (CPA)* will be used as a lexicographic technique to annotate and rank patterns in corpus samples. A follow-up to this pilot will consist of the description of more verbs and the combinatorics of core vocabulary nouns and adjectives.

2. Second language learners' intuitions about word combinations. To what extent are learners sensitive to lexical fixedness?

Irene Fioravanti¹, Alessandro Lenci², Anna Siyanova-Chanturia³

Speaker: Alessandro Lenci

¹Università per Stranieri di Siena, ²Università di Pisa, ³Victoria University of Wellington

Monday June 18th Session 1: 09.55-10.20

The present research focuses on the acquisition of verb-noun word combinations by second language learners of Italian. The phraseological approach distinguishes free combinations, collocations and idioms in terms of lexical restriction. Overall, research suggests that idioms and free combinations are less problematic for learners compared to collocations. Might this be due to their nature? Does the degree of restriction play a role in learners' vocabulary use? Do learners perceive the three types of word combinations differently? How likely do learners select a synonymic variants of word combinations? Two studies have been carried out to answer these questions. In Study 1, 120 Italian verb-noun word combinations were embedded in sentential context; an identical set of 120 sentences was further used in which the verb of the target combination was substituted with a synonym. English learners of L2 Italian were asked to rate acceptability of the sentences. In Study 2, the same materials were used.

3. Native and learner judgments of idiom properties: what do they tell us about L2 idiom knowledge?

Ferdy Hubers, Catia Cucchiarini, Helmer Strik

Speaker: *Ferdy Hubers*

CLST/CLS, Radboud University, the Netherlands

Monday June 18th Session 1: 10.20-10.45

Research has shown that L1 and L2 idiom processing is affected by idiom properties such as frequency, familiarity, transparency, imageability, and L1-L2 similarity (Steinel et al. 2007; Garcia, Cieslicka & Heredia, 2015).

Data on these idiom properties are usually obtained by asking people to indicate their intuitions about the idiom properties through subjective judgment scales (Bonin et al. 2013; Libben & Titone, 2008; Nordmann et al. 2014; 2017; Tabossi et al. 2011). When studying L1 idiom processing, it is obvious that the judgments should be collected from native speakers of the language under study. However, in research on L2 idiom processing and learning, it is not immediately clear from which raters such judgments should be collected. On the one hand, it seems reasonable to take the intuitions of native speakers as the benchmark, because native speakers can be considered as the model the L2 learners are trying to achieve. On the other, it is conceivable that native judgments are not in line with the intuitions and impressions of L2 learners and may have a biasing effect on the results. Collecting subjective judgments directly from L2 learners would then seem to be preferable. Relevant questions in this respect are whether L2 learners are at all capable of developing accurate intuitions about idiom properties and whether their intuitions are a better reflection of L2 idiom processing than the intuitions by native speakers.

In this talk we report on a study that addressed these research questions. Our results show that L2 learners provided reliable judgments of idiom properties and that these were more strongly related to their objective knowledge of idioms than the judgments provided by native speakers. When studying L2 idiom processing and learning it therefore seems preferable to take L2 learner intuitions as a benchmark rather than intuitions by native speakers.

4. What an individual difference approach may reveal about spoken idiom comprehension

Speaker: *Cristina Cacciari*

University of Modena-Reggio Emilia, Italy

Monday June 18th Session 2: 11.20-11.45

Idioms still represent a challenge to theories of language processing after more than 40 years from the first psycholinguistic study on idiom comprehension (Bobrow & Bell, 1973). Despite the fact that several models have been proposed to account for idiom processing, the complex cognitive architecture that stands behind the relatively easy and fast comprehension of familiar idioms has not yet been singled out in full details. In this talk, I present new evidence coming from an exploratory study about whether and to what extent individual differences in cognitive and personality variables are associated with spoken idiom comprehension in context. Language unimpaired participants were enrolled in a cross-modal lexical decision study in which semantically ambiguous Italian idioms predictable or unpredictable before the string offset, were embedded in idiom-biasing contexts. To explore the contributions of different cognitive and personality components, participants also completed a series of tests respectively assessing general speed, inhibitory control, short-term and working memory, cognitive flexibility, crystallized and fluid intelligence, and personality. Stepwise regression analyses revealed that online idiom comprehension was associated with the participants' working memory, inhibitory control and crystallized verbal intelligence, an association modulated by idiom type. Also personality-related variables (State Anxiety and Openness to Experience) were associated with idiom comprehension. These results contribute to the renewed interest on how individual variability modulates language comprehension, and for the first time document contributions of individual variability on lexicalized, high frequency multi-word expressions as idioms adding new knowledge to the existing evidence on metaphor and sarcasm.

5. Children are more likely to store internally coherent rather than frequent undersegmented chunks: Quantitative evidence from a corpus study

Speaker: *Robert Grimm*

University of Antwerp, Belgium

Monday June 18th Session 2: 11.45-12.10

One of the tasks faced by young children is the segmentation of a continuous stream of speech into discrete linguistic units. A possible strategy for bootstrapping the segmentation process is the wholesale storage of syllable chunks within an early proto-lexicon. Stored chunks can then be compared to one another and to incoming input in order to identify units such as phonemes and morphemes. Here, we investigate what types of chunks are most likely to be stored by children. Our method involves sampling syllabified utterances from transcribed child-directed speech as units that children could potentially store as undersegmented chunks. We vary sampled utterances according to (a) their frequency in child-directed speech and (b) the mutual predictability of their syllables (coherence). We then ask whether it is easier to identify single-word chunks on the basis of (i) their frequency or (ii) their coherence. Following this, we attempt to ascertain whether children are more likely to store coherent or frequent chunks. We argue that comparing chunks to one another and to incoming speech should quickly result in identifying words which are contained in a large number of stored chunks. Consequently, we consider it more probable that children store chunk samples that perform well at predicting when their constituent words are learned, and we consider it less likely that children store poorly performing samples. Our results show that especially coherent chunks are more likely to correspond to single words, and that coherent chunks samples consistently perform better at predicting when their component words are learned. These patterns suggest that children's early proto-lexica are more likely to contain coherent rather than frequent undersegmented chunks.

6. Learning Formulaic Sequences from Audio-Visual Input

Elke Peters, Eva Puimège

Speaker: *Elke Peters*

University of Leuven, Belgium

Monday June 18th Session 3.1: 13.30-13.55

There is some evidence that formulaic sequences (FS) can be learned incidentally through reading (Pellicer-Sánchez, 2015; Peters, 2012) and reading-while listening (Webb et al., 2013). However, little is known about learning FS from audio-visual input. This is surprising given that corpus research has shown that the oral discourse typical of audio-visual input could provide sufficient exposure to FS for incidental learning to occur (Lin, 2014).

The present study explores the incidental learning of FS from audio-visual input. A pretest-posttest, within-subject design was adopted. English-as-a-foreign-language learners (L1=Dutch; N = 42) watched a one-hour English language documentary without subtitles. The results showed significant learning gains at the level of form recall and meaning recall as well as the incremental nature of learning FS (from meaning knowledge to form knowledge). Our findings will be discussed in light of factors that might affect learning gains (learners' vocabulary size, type of FS, partial knowledge).

7. Learning Dutch L2 idioms: the role of practice intensity, practice modality and idiom properties

Catia Cucchiarini, Ferdy Hubers, Helmer Strik

Speaker: *Catia Cucchiarini*

CLST, Radboud University, the Netherlands

Monday June 18th Session 3.1: 13.55-14.20

Idiomatic expressions are a particular category of formulaic language that turn out to be especially problematic for L2 learners, both in language comprehension and in language production (Cieślicka, 2006; Concklin & Schmitt, 2008; Steinel et al. 2007; Wray, 2000). However, idioms are an essential component of vocabulary and L2 learners indicate that learning idioms is important (Liontas, 2002). The majority of studies so far address L2 English idiom learning, while less is known about learning idioms in other L2s. Previous research has shown that increased exposure and practice facilitate the acquisition of idiomatic language, but it is not completely clear how practice can best be provided and how the frequency and the nature of practice interact with important idiom properties such as transparency and cross-language overlap.

In this presentation we report on research that was conducted within the project Idiomatic Second Language Acquisition (ISLA), where L2 idiom learning is studied in relation to L1 and L2 idiom comprehension and production.

We report on a study based on a within-subject pre-test post-test design in which 42 German learners of Dutch L2 practiced with 60 Dutch idiomatic expressions. A Computer Assisted Language Learning (CALL) system was employed to present different types of exercises focusing on meaning and form with different frequency and in different modalities. We discuss the effects of intensity of practice on Dutch L2 idiom learning in relation to modality of practice and idiom properties such as transparency and cross-language overlap.

8. The role of input and cross-linguistic overlap in L2 formulaic language acquisition and processing

Speaker: *Kathy Conklin*

University of Nottingham, United Kingdom

Monday June 18th Session 3.2: 14.40-15.05

Formulaic language can represent a challenge to even the most proficient L2 learners. Evidence is mixed as to whether native and non-native speakers process it in a fundamentally different way, whether exposure can lead to more nativelike processing for non-natives, and how L1 knowledge may benefit comprehension. Across several studies, we investigate how non-native speakers with various levels of proficiency process idioms encountered in their L2. Generally, results support the view that L1 knowledge is automatically activated from the earliest stages of processing, regardless of whether sequences are congruent, and that exposure and advanced proficiency can lead to nativelike formulaic processing in the L2. In a final study, look more explicitly at the role of input in formulaic language processing in very early learners. We explore the processing of formulaic sequences that are present in the textbooks of beginner learners of English in Japanese secondary school. Students responded to formulaic that occurred in their textbooks significantly faster and more accurately compared to formulaic language that is frequent in natural language, and they were sensitive to the frequency of occurrence in their textbooks. This has important implications for teaching and material design for formulaic language.

9. Idiom acquisition across the lifespan

Speaker: *Simone Sprenger*

University of Groningen, the Netherlands

Monday June 18th Session 3.2: 15.05-15.30

Observations in the L2 classroom, as well L2 acquisition research consistently show that for L2 learners the acquisition of multi-word expressions is a challenge. That is, learners are less likely to know idioms ("he spilled the beans"), collocations ("heavy rain"), or even linking adverbials ("in other words"), let alone use them with the same naturalness as a native speaker. This is a matter of concern, because it limits the learner's ability to express herself fluently in a way that is readily understood (Pawley & Syder, 1983). The L2 learners' lack of proficiency with respect to such fixed phrases is often contrasted with the supposedly full proficiency of native speakers, who are assumed to have easy access to a large repertoire of figurative and non-figurative expressions. However, it is unclear how this proficiency develops and how much variability there is among native speakers. Here we show that the shape of the acquisition curve is for idioms across the life span is highly similar to that for single words, but that there is a clear lag in acquisition: idiom knowledge in native speakers of Dutch only shows a sharp increase after adolescence, at a point in time when the rate of single word vocabulary acquisition is well beyond its highest peak. That is, contrary to what is often implicitly assumed in the literature, idiom acquisition is not only slow in L2 learners, but in native speakers as well. Our results imply that the hypothetical baseline of the native speaker as a proficient user of idioms only holds for speakers above thirty years of age, and thus does not apply to the average (student) participant in controlled experiments.

10. The dynamics of understanding literal and non-literal meaning in idiomatic expressions: A look at language learners

Speakers: *Sara Beck and Andrea Weber*
Eberhard Karls Universität Tübingen, Germany
Tuesday June 19th Session 4.1: 09.00-09.25

Our research project on figurative meaning constitution investigates how language learners understand idiomatic expressions. We are specifically interested in the competition between immediate compositional processes and non-compositional meaning and the conditions in which one might be prioritized over the other. We do so by researching the flexibility of the involved processes and their role during the course of language development. Considering both language acquisition by L1 children and L2 adults, we are looking for developmental changes particularly when comparing access to the idiomatic “chunk” and access to the individual constituent words. In our presentation, we will give a short overview of our current studies before focusing on a series of cross-modal priming experiments in English. In these studies, L1 (American English) and L2 (German) adults listened to neutral sentences containing idioms (i.e., John likes to pull my leg.) and responded to a lexical decision task in which responses were compared between words related to the overall meaning of the idiom (JOKE), the literal meaning of an individual constituent word (WALK), or a control word (MILK/SHIP). Individual experiments varied in factors such as the translatability of idioms between German and English, the global experimental context, and the timing of the target presentation. While we find that highly proficient speakers of English generally behave like native speakers, the limits of these capabilities will be discussed.

11. Native and Bilingual Idiom Comprehension: Effects of Word Frequency, Idiom Properties and Language Status

Wendy van Ginkel, Ton Dijkstra

Speaker: *Wendy van Ginkel*

DCC, Radboud University, the Netherlands

Tuesday June 19th Session 4.1: 09.25-09.50

Idiomatic expressions such as *to kick the bucket* have their own specific meaning as a whole (*to die*), but still consist of a set of literal words with their own respective meaning. The meaning of some expressions is more transparent, and as such can be more readily derived from its literal word constituents (e.g. *to hit two birds with one stone* means *to do two things at once*), whereas the meaning of others such as *to kick the bucket* is more opaque and harder or impossible to derive from its constituent words. Furthermore, some idioms such as the previous examples are plausible as a completely literal sentence, whereas others are not literally plausible (e.g. *to be on cloud nine*). Models of idiom comprehension in the first (L1) and second language (L2) differ in the extent to which they allow for activation of the literal word meanings of constituent words within an idiom when that idiom is being processed. I will present data from a L1/L2 Dutch lexical decision study examining the activation of figurative and literal meaning aspects of idioms as a whole as well as their constituent literal words. Both L1 and L2 (L1 German) advanced speakers of Dutch showed similar priming for figurative and literal meaning aspects of idioms, suggesting similar processing mechanisms for these idioms in general for the language groups. I will further present data showing that word frequency of literal words within an idiom can affect processing of lexical decision targets in the L1 group. Idiom transparency and literal plausibility affected processing of targets in the L1 group, but not in the L2 group. Findings will be discussed in terms of models of idiom comprehension, and differences between the language groups will be discussed.

12. Electrophysiological markers of non-native novel and conventional metaphor processing

Speaker: *Karolina Rataj*

Adam Mickiewicz University, Poland

Tuesday June 19th Session 4.2: 10.05-10.30

Electrophysiological studies on metaphor processing in the native language have revealed modulations of the N400 component by metaphor conventionality, with smaller N400 amplitudes evoked by conventional than novel metaphors. The late positive complex (LPC) has shown less consistent results, with both reduced and increased amplitudes observed for metaphoric utterances. Recent studies have, however, demonstrated the sensitivity of the LPC to metaphor conventionality, reflected in smaller amplitudes evoked by novel than conventional metaphors. One explanation of this effects is that in the case of novel metaphors, sustained negativity reflecting continued retrieval of semantic information overlaps with the LPC, thus reducing its amplitudes. Since conventional metaphors in the non-native language are not conventionalised to the same degree as in the native language, it can be predicted that the differences in the N400 and LPC amplitudes between conventional and novel metaphors in the non-native language should be less pronounced than in the native language. An event-related potential (ERP) study discussed here examined this hypothesis. Participants performed a semantic judgment task in response to literal, conventional metaphoric, novel metaphoric, and anomalous word pairs in their native (Polish) and non-native (English) language. The results showed a general language effect in the N400 time window, with an overall reduction of the N400 amplitudes for the non-native language. At the same time, conventional metaphoric utterances evoked overall smaller N400 amplitudes than novel metaphoric word pairs. An intriguing pattern was found in the LPC time window. While in the native language LPC amplitudes evoked by novel metaphoric utterances were significantly reduced compared to conventional metaphors, in the non-native language LPC amplitudes for conventional and novel metaphoric word pairs converged. This result indicates continued retrieval of semantic information and increased semantic integration demands in conventional metaphor processing in the non-native language.

13. L2 idiom processing: Combining psycholinguistics and applied linguistics

Ferdy Hubers, Catia Cucchiarini, Helmer Strik, Ton Dijkstra

Speaker: *Ferdy Hubers*

CLS/CLST, Radboud University, the Netherlands

Tuesday June 19th Session 4.2: 10.30-10.55

Idiomatic expressions are known to be particularly challenging even for highly proficient L2 learners (Ellis, Simpson-Vlach, & Maynard, 2008; Pawley & Syder, 1983). Psycholinguistics and applied linguistics have both investigated the comprehension and production of idiomatic expressions by L2 learners from different perspectives, by using different techniques, and often obtaining different findings. In psycholinguistics, research mainly employs experimental techniques that directly tap into the processing of L2 idiomatic expressions to investigate the underlying processes and representations (Siyanova-Chanturia, 2015). Less attention is usually paid to whether the non-native participants really know the idiomatic expressions investigated. Applied linguists, on the other hand, usually analyse learning gains and focus on the end product (responses correct).

Our research combines psycholinguistics and applied linguistics, by investigating how L2 learners who practised with Dutch idiomatic expressions through a Computer-Assisted Language Learning (CALL) system, performed in a psycholinguistic word-naming study that tapped into the actual processing of L2. The combination of controlled training and experimental testing contributes to bridging the gap between psycholinguistic and applied linguistic approaches.

14. A dynamical systems approach to language modelling

Speaker: *Rüdiger Thul*

University of Nottingham, United Kingdom

Tuesday June 19th Session 5: 11.30-11.55

Numerous studies have demonstrated a processing advantage for formulaic language compared to novel or grammatically generated language, e.g. the idiom “break the ice” compared to “crack the ice”. Crucially, it appears that this processing advantage cannot be completely attributed to the predictability of upcoming words in the formulaic language. This has led us to the conclusion that what underpins the processing advantage, at least for certain phenomena like idioms, reflects something above and beyond predictability. However, the analysis techniques we generally apply do not allow us to explore what this might be.

In this talk, I will present a novel method for modelling the reading patterns elicited by idioms. The model is based on reading times from an eye-tracking study comparing idioms and controls embedded in sentences when they were in their canonical form (“break/crack the ice”) and when 0-4 additional word(s) intervened before the expected completion (“break/crack the amazingly cold ice”). In our model, we attribute “linguistic forces” to words, and these forces can either increase or decrease the reading times. Translating this concept into the language of dynamical systems, we found that the reading time of the final noun, R_f , is captured by the following:

$$R_f = R_0 \exp\{-\eta T^\alpha R_0^{\alpha\eta}\}.$$

Here, R_0 is the reading time of the noun in the canonical form of the idiom, T corresponds to a constant that characterizes how “strongly” the added word(s) move the noun, α is a nonlinear cooperativity factor, and $\eta = \pm 1$ indicates whether the additional word(s) decrease or increase reading time. Such an approach allows us to identify potential mechanisms that give rise to the observed reading patterns and goes beyond a statistical description.

15. Computational modelling of formulaic expressions

Louis ten Bosch, Alexander Wahl, Ton Dijkstra

Speaker: *Louis ten Bosch*

Radboud University, the Netherlands

Tuesday June 19th Session 5: 11.55 – 12.20

In this presentation, we discuss a computational model for left-to-right sentence processing that takes the unfolding word sequence as input, and outputs the probability of literal and non-literal interpretation as a function of time. Like other models, such as the computational model of word sequence acquisition by McCauley and Christiansen (2014), our model is based on the use of word-word, word-concept and concept-concept transition. The aim of the model is to better understand underlying cognitive processes by simulating latency effects as function including frequency, transparency, and contextual diversity (McDonald and Shillcock, 2001; Hills et al., 2010), by focusing on prediction of reaction times and words as a function of predictors including idiom frequency, semantically opaqueness and compositionality.

Our model, the Figurative and Literal Interpretations in Sentence Processing ('FLISP') model, updates several different hypotheses in parallel. An input sentence like 'John kicks the ball' allows an interpretation that can be derived from the meaning of all the individual constituent words appearing in the sentence while 'John kicked the bucket' is likely to activate a figurative interpretation based on the joint presence of words in a particular order. The processing of word sequences narrowly interacts with the processing of individual words (see, e.g., Grimm et al., 2017, and references therein).

The FLISP model will be presented in the context of a few alternatives such as graphical models, and models based on dynamical system theory. We assign conditional probabilities to concepts on the basis of input forms observed thus far as well as for forms that are necessary for a concept to be attested ("concept forms"). Multiple one-to-many mappings from forms to meanings are hypothesized and dynamically ranked over time, but not reverse: a single meaning cannot be mapped to multiple forms.

ACKNOWLEDGEMENTS

This workshop is part of the research program Free Competition in the Humanities with project number 360-70-510, which is financed by the Netherlands Organisation for Scientific Research (NWO). We would also like to thank the International Office of Radboud University for supporting this workshop.