This paper addresses approaches to verb argument structure from the point of view of the information which can be assumed to be lexically encoded in the verb. It explores ways in which speakers’ sensitivity to verbs can be investigated experimentally across types of expressions, including idioms/non-literal language, and reports findings from recent empirical research in that domain.

Keywords: lexically encoded information, verbs, argument structure, sensitivity, idiom processing

1. Why verb argument structure?

The argument structure of verbs has been in the focus of extant theoretical proposals on the nature of lexical knowledge and its interface with syntax. Depending on the tradition and framework, various terms have been used to cover what is essentially versions of the same phenomenon: verb subcategorization, theta-grid, verb valence, lexical specification. Discussions and disagreements have typically centered around the question of the lexicon and the extent to which lexical information is inherently linked to the workings of grammar or the two are dissociated and function independently. Crucially, the question is whether lexically encoded information, and thereby, verb arguments, pre-specify the structure of sentences headed by this verb. In addressing this question, Jackendoff has proposed a model where all three levels of language structure (phonology, syntax and semantics) both enjoy certain independence, but at the same time richly interface with each other via interface rules (Jackendoff, 2002). The question is reversed to inquire what aspects of language need to be stored in long-term memory by virtual necessity and what aspects do not. This reversal of focus serves to anchor models of language in the domain of language learning and processing and explain language structure from the point of view of how humans can possibly acquire, understand and produce language via basic cognitive mechanisms assumed to support language (Pinker & Jackendoff, 2005). Jackendoff thus suggests a radical view of the word as distributed between the three structures (phonology, grammar and semantics) and represented
as a multiple linking rule where a phonological representation is linked to grammar information (morphological structure and word category) and to semantics. “Thus, a word is a stored association of three kinds of structures, plus a subscript that links them. In other words, the lexicon is not a component that inputs to the syntactic component, but rather a part of the interface components: it helps establish the relation among phonology, syntax, and semantics.” (Jackendoff, 2002, p. 27–28).

Other models within the same tradition typically include the most salient aspects of words necessary both for learning words, their long-term storage and active use (e.g., lexical access). Common to these models is that, for verbs, the nature and number of arguments they take will be part of the semantic representation of the verb. Thus, in Levelt’s (1993) influential model, arguments are part of the meaning section of the lemma and are conceived as the arguments of conceptual functions. A related, albeit distinct proposal, the Lexical Quality Hypothesis, views lexical items as the nexus of phonological, orthographic and semantic information (Perfetti & Hart, 2002; Perfetti, 2007). Multiple encounters with the word during the process of word learning ensure the creation of this common core (lexical) representation. Fast and efficient retrieval of a word relies on the quality of all the features which form part of the word representation. Thus, upon hearing a word (and its auditory signature), components of its meaning will be simultaneously activated, including activation of the number and types of arguments a verb may allow for (Koenig et al, 2003). Importantly, on the proposal by Koenig and colleagues, lexically encoded information is indifferent to whether the arguments need to be overtly realized at the level of syntax or not. This essentially implies that certain aspects of argument information are only obligatorily represented, but need not be overt. This information underlies native speaker judgements of acceptability, as well as expectations and predictions of upcoming material in language processing. It can be further stipulated that, in such a way, lexical information both enriches processing by making available free (thematic) slots, and constrains language structure/grammar by constraining the number and nature of these slots (see Dimitrova-Vulchanova, 1999 for a detailed discussion and a theoretical model; Dimitrova-Vulchanova & Weisgerber, 2007).

2. What does sensitivity to argument structure indicate?

Given that argument structure plays a central role in speakers’ word representations and is activated upon encountering the word, we can expect that the information about the number and nature of arguments a verb encodes reflects native speakers’ language competence, and can, as such, be used in experimental studies of language as a proxy for language skills. Since fast and efficient retrieval of a word relies on the quality of all the features which form part of the word representation, in L1 speakers, sensitivity to verb argument structure
would indicate adequate quality of lexical storage and lexical integration of the verb in the lexical network of the mental lexicon.

On an assumption that word learning relies on similar cognitive mechanisms and leads to similar results in second language users, sensitivity to verb argument structure can be used as a litmus of lexical consolidation and integration also in that population. However, L1 and L2 speakers differ on one important parameter, namely that lexical knowledge in the L1 tends to interfere with the acquisition of new words in the second language. It has been suggested that L2 users “suffer” from a full transfer of lexical content from the L1 onto equivalent words in the L2, a phenomenon described by the *Full Transfer Hypothesis* (Stringer, 2010). With time, and more exposure to L2 input, it becomes possible for L2 speakers to adjust their lexical representations, and “re-conceptualise” the situations denoted by L2 verbs (Treffers-Daller & Calude, 2015). Extant research suggests that sensitivity to verb argument structure can reveal the level of proficiency in second language learners and that differences between beginner-level and more advanced learners can be observed in that domain (Larrañaga, Treffers-Daller, Tidball & Ortega, 2012; Treffers-Daller & Calude, 2015; Treffers-Daller & Tidball, 2015; Johnsen, 2016; Reine, 2016). In particular, those studies suggest that, at initial stages of second language learning, L2 users tend to be influenced by transfer from the native language of the ways in which verbs conflate specific semantic features, including the encoding of arguments, and how participants in the situations these verbs denote are lexicalized in the L1. This research also suggests that more advanced L2 proficiency and frequent exposure to the L2 can lead to escaping the transfer trap (Cabrera & Zubizarreta, 2005; Stringer, 2010).

3. Studying sensitivity in L1 and L2

There are different ways of studying sensitivity to argument structure experimentally. The L2 studies mentioned above have used free production tasks where L2 learners are asked to produce narratives and their production is coded for analyses. What is of interest here is the extent to which verbs are used in the expected target language syntactic context. Comparisons are then made with production in the same task by native speakers of the language. An alternative strategy is to track L2 learners’ behaviour when listening to target language stimuli sentences involving the verbs of interest. In this type of design it is possible to assess the extent to which L2 users successfully expect and predict what noun phrase will fill a specific argument slot in the sentence as it unfolds in time. Using this design, and consistent with the Visual Word Paradigm (VWP) (Allopenna et al., 1998; Altmann, 1997; Altmann & Kamide, 1999), we approached the question of whether we would find argument prediction differences between less proficient and advanced Norwegian learners of English as a second language.
**Method**

Participants in the study were 25 16 year-old high school students (intermediate proficiency group) and 26 university students (age range 20-28, MA 24.07). Their task was to listen to English sentences while their looks towards visual stimuli presented on a computer screen were being tracked. The stimuli included a picture matching the target noun phrase filler, a phonological cohort (a word with the same phonological onset), and two distractor images (cf. Fig. 1). Data were collected with a Tobii T120 eye-tracker, at a sampling rate of 60 Hz. In the preparatory phase we collected data from 105 L1 speakers of English (age range 18-45; MA 19.75) using a cloze test with 77 frequent English verbs which were expected to be familiar to the L2 speakers. The native speakers completed the study via an online form. Cloze tests are employed to assess the probability of specific words following a target word of interest. Thus, in the sentence “John kicked the ----” the interest is in what NP is most likely to occur in the position immediately following the verb *kick*. Depending on the percentage of native speaker responses that converge on a specific filler, the cloze probability of possible fillers of that position can be established. In the case of *kick*, both data from native speaker elicitations and corpora suggest that “ball” is the most common filler. This essentially entails that, in the context of a sentence headed by *kick*, native speakers of English will look faster to a picture of a ball over pictures representing any other object, suggestive of anticipation of the NP “the ball” in the direct object position of that verb.

![Figure 1](image.png)

**Figure 1.** Example of stimulus display for sentence “Jesse fastened the belt eagerly.” With a picture of the target filler (*belt*), of a phonological competitor (*bell*) and two distractor images (*flowers* and *flour bags*).
For the elicitation from native speakers we selected thirty-five (35) verbs that were defined as constrained in their lexical selection of NP fillers (e.g., water, milk, squeeze, stroke). Twenty-one (21) verbs were defined as moderately constrained (e.g., paint, load), while additional twenty-one (21) verbs were defined as open in their direct object selection (e.g., play, remove). Following the data elicitation procedure with native speakers, the target verbs for the L2 eye-tracking study were divided into 3 categories. For each verb, if three unique answers or fewer collectively constituted at least 55% of all the given answers, the verb was assigned to the most constrained category. If four to six unique answers constituted at least 55% of all the answers, the verb was assigned to the moderately constrained category. If seven or more unique answers were needed to constitute 55% of all answers, the verb fell into the least constrained category. When judging the uniqueness of an answer, only heads of noun phrases were counted, ignoring any prepositional phrases or adverbial phrases. The stimuli verbs for the experiment were also selected with regard to depictability and possible phonological cohorts. We also included a base-line neutral category of verbs like choose, pick, select, take which place no specification requirements on their direct objects. The final selection included 16 verbs in each category. Each stimulus sentence was five words long, and consisted of a subject [AGENT], verb, definite article, object noun, and an adverb. As subject of the sentences, the five names Alex, Sam, Charlie, Jesse and Mary were used. An adverb was included as the final word of each utterance to avoid any unwanted effects from having the target noun as the utterance-final word. From the data, 2.03% data were excluded from analyses (these were trials with less than 25% registered gaze points). Prior to participation in the experimental study, participants completed online tests of L2 (English) grammar proficiency (https://www.examenglish.com/leveltest/grammar_level_test.htm) and vocabulary size (http://vocabulary.ugent.be/). Scores from these tests were included in the analyses as predictor variables of performance on the eye-tracking test. Statistical analyses on the gaze data included an analysis of variance (ANOVA) for effect of verb category and an analysis of covariance (ANCOVA) for the effect of language proficiency measures (grammar and vocabulary) on performance.

Results

Anticipatory looks towards the target item reflect an understanding of the verb and an ability to map that verb onto a likely situation representing the referent of the verb argument(s) (Brock et al., 2008). Figure 2. shows the development of gaze proportion towards the target object from the onset of the verb for the less proficient group of participants, while Figure 3. shows the proportion of anticipatory looks towards the target object for the advanced proficiency group.
Figure 2. The X-axis shows gaze proportions for the target item in the target-present condition. The Y-axis shows time in milliseconds relative to verb onset. VerbCat 0 = baseline, 1 = least constraining, 2 = moderately constraining, 3 = most constraining. The solid vertical reference lines at 400 and 1400 ms denote the starting point and end point of the critical time window. The dashed vertical reference line denotes mean object onset time.

For the less proficient L2 speakers, the mean onset time for looks towards the target was 651 ms indicative of anticipation of the direct object. Increasing gaze proportions for the target item in Fig. 2 displayed a larger effect for the more constraining verbs than the less constraining ones, and this larger effect lasted until around 1400 ms post verb onset. A post-hoc pairwise comparison of verb categories was applied to determine whether the difference in gaze proportions across verb categories were significant. The significance was adjusted for multiple comparisons with Bonferroni correction. Gaze proportions for verb category 3 (the most constraining category) were significantly larger than proportions for both category 0 (the baseline category) and 1 (the least constraining of the test verb categories). There was no significant difference between categories 2 and 3. In other words, the only verb category that led to significantly increased looks to the target object compared to other categories was the most constraining one, and only compared to the least constraining of the three categories as well as the baseline category.
Figure 3. Mean proportions of looks towards the target across all verb sentence types. The dotted line indicates the critical time window.

The post-hoc pairwise comparison of proportion of looks towards the target in the advanced proficiency group revealed that the most constrained verb category elicited significantly more anticipatory looks and was thus significantly different from all other verb types. In addition, the other verb types did not differ from each other concerning anticipation of the direct object filler. While these results appear similar to the ones obtained for the less proficient group, there are differences as well. In the advanced proficiency group, looks towards the target start around 600 ms after verb onset, while for the less proficient group proportion of looks starts increasing only around 700 ms after verb onset. Also numerically, the mean proportion of looks towards the target for the most constraining category of verbs during the critical time window (400 – 1400 ms) is significantly different between the less proficient group (.49112) and the advanced proficiency group (.60423) ($p < .01$).

The analysis of covariance (ANCOVA) in the less proficient group revealed a significant within-subject effect of grammar score, $F(3,66) = 3.202$, $p = .029$, $\eta^2 = .127$, but no significant effect of vocabulary score, $F(3,66) = 1.516$, $p = .218$, $\eta^2 = .064$. No significant between-subjects effect was found in that group (see Fig. 4). The analysis of covariance (ANCOVA) for the advanced proficiency group revealed a between-subjects effect of vocabulary knowledge approaching significance ($F(3.871), p = 0.61, \eta^2 = .139$) (Fig. 5). Thus, participants with higher vocabulary scores displayed faster and more looks to the target in all verb categories in the advanced proficiency group, while in the less proficiency group the effect of verb category was visible in gaze behaviour only for participants with higher grammar scores.
Figure 4. Relationship between gaze proportions for target-and results in grammar test less proficient group. Only for participants scoring C1 or C2 in the grammar test does the pattern of increased looks towards targets in verb categories 2 and 3 compared to 0 and 1 emerge.

Figure 5. Relationship between gaze proportions for target-and results in the vocabulary test advanced proficiency group. Participants with higher vocabulary scores look longer to target for all verb categories.
What this study demonstrates is that different language skills predict anticipatory gaze behaviour at different levels of proficiency. Thus, for the advanced L2 learners, vocabulary predicts expectations on what noun phrase is the most appropriate filler of the internal argument position of each and every verb in all categories used in the experiment, while for the less proficient group, grammar competence is a stronger predictor of sensitivity, and only for the most constrained category of verbs. This suggests that the less proficient L2 learners display sensitivity primarily for verbs with more constrained argument frames where only a small number of NP fillers can be used felicitously, while for the advanced learners, sensitivity, as reflected in anticipatory gaze behaviour and the effect of vocabulary knowledge, applies to all verb types regardless of argument selection. These results provide evidence of across-the-board sensitivity to verb argument structure in more highly proficient learners against more modest anticipatory performance in the less proficient group and only for highly selective (constrained) verbs. Further support of this difference is indicated in the time-line of gaze behaviour where looks towards the target are registered in an earlier time-window after verb onset for the advanced group in comparison to the less proficient L2 user group. Finally, the less proficient group display less distinct gaze behaviour as reflected in the smaller difference in slope between constrained and moderately constrained verbs.

4. Formulaic language. Idioms

Formulaic language refers to sequences of words that occur frequently in language. Examples are fish and chips, don’t worry about it, take the wheel. Typically, these are multi-word expressions used in every-day interactions characterized by their fixed grammatical form (Kuiper, 2017). According to estimates, multi-word expressions constitute one-third to one-half of daily language routines (Conklin & Schmitt, 2012) and are assumed to aid processing by reducing the load on working memory. Multi-word chunks have also been proposed to play a central role in communication and language evolution in that they help overcome the Now-or-Neve r bottleneck constraint on fast and efficient communication (Christiansen & Chater, 2016; Isbilen & Christiansen, 2020). The interpretation of multi-word expressions (also labelled chunks) is (largely) non-compositional, meaning that the overall meaning of the expression cannot be computed online and directly as a function of the meanings of its constituents. Formulaic language covers a number of different categories, idioms, metaphors, clichés, phrasal units. The extent to which interpretation is non-literal depends on the type of expression. Among these, idioms are a central category which has attracted attention both in language theory and experimental studies of language due to their characteristics.

Different accounts have been proposed to explain the nature of idiomatic expressions and the ways in which they are processed. A central question addressed in all approaches is whether the figurative (non-literal) meaning of the
expression is accessed directly in the mental lexicon, much in the way speakers retrieve word meanings or there is an initial step involving a literal interpretation of its constituents, as well as in what order these operations are computed. An approach which assumes a single step processing is for instance the Direct Access model whereby there is no need for initial activation of the literal meaning(s) (Gibbs, 1994) and the figurative meaning is retrieved directly using contextual cues. In contrast, the Standard Pragmatic approach assumes a two-step process with initial activation of the literal meanings associated with the constituent words in the expression, only arriving at the intended figurative non-literal interpretation at second step (Grice, 1975). There are also accounts which suggest simultaneous processing. For instance, on the Lexical Representation Hypothesis (Swinney & Cutler, 1979), idioms are assumed to be stored as lexical items which can be retrieved fast, at the same time engaging in a second parallel process of lexical decomposition. There is consensus in current thinking that idioms are complex expressions and their processing cannot be reduced to either direct lexical access or literal processing, and that interpretation depends on a number of factors (Vulchanova et al., 2015; Vulchanova et al., 2019a). For instance, according to the Configuration Hypothesis idioms are represented in a distributed way and processed as complex expressions, much like other instances of similar syntactic complexity (Cacciari & Tabossi, 1988), while the hybrid approach (Titone & Connine, 1999) attributes the variation in idiom processing to the degree of decomposability of the expression as the most important criterion.

In a series of recent studies we provide evidence that the literal meanings of idioms are often activated (Vulchanova et al., 2019b; Chahboun et al., 2016; Milburn, Vulchanov & Vulchanova, 2021) and that they often compete with the target figurative interpretation of the expression. Whether they are by-passed or not depends on the decomposability of the expression, the individual variation in level of familiarity (size of mental lexicon), developmental stage (child or adult user) and level of proficiency, e.g., whether the participant is an L1 or an L2 speaker.

Idioms take time to be acquired in L1 despite suggestions and evidence that children initially acquire (literal) language chunk-wise (Vulchanova, Vulchanov & Stankova, 2011; Isbilen & Christiansen, 2020). This is in contrast to words, which are among the first language units to be acquired and used by infants (Vulchanova & Vulchanov, 2021). This may indicate that the conditions that ensure successful word learning and idiom competence are not identical. It has also been documented that adult L2 learners have difficulty with formulaic language, that they produce fewer formulaic sequences than native speakers and that they may be less able to use multiword units when identified. It appears that the greatest problem for L2 learners is the difficulty knowing when to process figuratively. This is evidenced by studies documenting that L2 speakers process idioms at a similar speed to novel literal phrases suggesting that these two types of expressions are processed similarly (Arnon & Christiansen, 2017; Conklin &
Schmitt, 2008, 2012; Siyanova-Chanturia et al., 2011a; Siyanova-Chanturia et al., 2011b). This is in contrast with faster idiom processing by native speakers. Now the question arises of what factors might support and enhance idiom processing in both L1 and L2 speakers. We identified two central factors of relevance. The presence of supportive context – which is an idiom-external factor – is typically acknowledged as scaffolding figurative interpretation. A second factor which might influence the extent to which the expression is processed with the intended figurative interpretation or literally is the collocational frequency of its constituents and the rate at which they co-occur in language. This is an idiom-internal factor which, in all likelihood, aids idiom recognition and retrieval.

The extent to which a literal computation competes with the figurative one can be predicted based on how often the idiomatic key (the word which determines the idiomatic meaning) co-occurs with the other constituents of the idiom in comparison to co-occurrence with other words in free expression collocations. For verbs, this can be operationalised by the notion of *cloze probability* of the dependent constituent (Direct Object) of the verb. For instance, the verb *kick* takes an NP headed by “ball” as the most frequent filler of its complement position, with native speaker cloze probability of 54% and on-line corpora cloze probability of 14.58%. In contrast, the co-occurrence of *kick* with *bucket* as in the idiom *kick the bucket* is very low, albeit varying depending on corpus. Thus, it can be predicted that the two possible fillers will not compete, due to the cloze probabilities of each of those fillers in the context of *kick* (see Vulchanova et al., 2019a for a theoretical discussion).

In a series of on-going studies we addressed the question: Do native and non-native English speakers rely on different strategies—«word by word» processing vs «whole phrase» retrieval—when retrieving and producing idioms? The experimental design explored the role of the two main factors, the presence or absence of biasing context and the role of collocational frequency of the target word in the idiomatic expression contra highly frequent and low frequency words in the same position. We selected idioms comprising a verb head and a complement which allows for controlling the frequency of items in the complement position. The study design was based on the Visual World Paradigm (VWP) whereby participants hear sentences and their looks towards a visual display are being recorded. The visual display included stimuli referring to possible objects of the verb: the target figurative constituent, a high frequency competitor for the same position, a low frequency competitor and a lure. The auditory stimuli included common idiomatic expressions introduced either introduced by a figurative or literal biasing context or no context at all. Participants in the study were advanced L2 learners of English and native speakers of the language. The results for the L2 learners have already been reported in Milburn, Vulchanova & Vulchanov (2021), while the combined analyses of the data from the two groups are currently in preparation in Milburn et al. (in prepara-
tion). Overall, the prediction that L2 learners and native speakers of the language process idioms differently is being confirmed.

5. Closing remarks and implications

The overview of experimental evidence presented here suggests that sensitivity to language structure develops over time, both in the first and second language. It has also been argued and demonstrated that sensitivity to verb argument structure is a central component of language competence in both native and second languages. While many of the mechanisms on which language learning relies may be the same for both first and second languages (such as e.g., phonological and working memory, statistical learning, Vulchanova & Vulchanov, 2021), there remain differences. Processing depends on individual user characteristics and whether the user has native language competence or second language competence. Also, there is an abundance of evidence that children who are still acquiring the language and adult speakers process language differently. Thus, for idioms, child language users display a tendency for literal interpretation, which is typically more rarely observed in adults (Vulchanova et al., 2011; Chahboun et al., 2016).

With time processing becomes more automatized and adult L1 speakers more often rely on chunks, most probably in overcoming the bottlenecks of communication (Christiansen & Chater, 2016; Milburn et al., in preparation). Finally, it has been suggested that idioms are a useful lens for the study of sensitivity to the lexical properties of words, language processing, and can provide useful evidence also concerning the properties of literal language.

REFERENCES


Treffers-Daller, Tidball 2015: *Treffers-Daller, J., F. Tidball*. Can L2 learners learn new ways to conceptualise events? Evidence from motion event construal among English-speaking learners of French. – In: Schmitz, K., P. Gujarro-
Arguments for good or bad: Sensitivity to argument structure and idiom processing across ...


Prof. Mila Vulchanova, PhD
Valentin Vulchanov, PhD
Norwegian University of Science and Technology
Edvard Bulls veg 1, Trondheim 7491, Norway

Published: 5 July 2022