Helping learners engage with L2 words: the form-meaning fit

Julie Deconinck*, Frank Boers°* and June Eyckmans*°

*Free University of Brussels (VUB), Department of Languages and Literature, Pleinlaan 2, 1050 Brussels, Belgium.

°Erasmus University College, Department of Applied Linguistics, Pleinlaan 5, 1050 Brussels, Belgium.

Julie.deconinck@gmail.com +32-2-6293884

Frank.boers@ehb.be +32-2-6291074

June.Eyckmans@ehb.be +32-2-6291873

Engaging with L2 words: the form-meaning fit

Abstract

The pace at which new words are acquired is influenced by the degree of engagement with them on the part of the learner. Insights from Cognitive Linguistics into the non-arbitrary aspects of vocabulary can be turned into stimuli for such engagement. The majority of Cognitive Linguists' proposals for vocabulary teaching aim at helping learners appreciate the way a single word form can develop different meanings. This, however, presupposes knowledge of the 'basic' meaning of that word. We report an experiment in which learners under an experimental treatment were stimulated to consider the possibility that the form-meaning link in target words might not be fully arbitrary. The mnemonic effect of this task-induced engagement was assessed in relation to comparison treatments in immediate and delayed post-tests measuring both receptive and productive knowledge. Results show that simply prompting learners to evaluate the form-meaning match of words can foster vocabulary acquisition, although not all target words lends themselves equally well to this type of engagement.

Introduction

The chances of an L2 word becoming entrenched in long-term memory are influenced by the degree of cognitive (and affective) involvement on the part of the learner (e.g. Laufer and Hulstijn 2001). Schmitt (2008: 339-340) uses "engagement" as an umbrella term to refer to any activity on the part of the learner that involves more attention to or manipulation of lexical items. Others (e.g. Barcroft 2002) have adopted the term "elaboration" to describe the various mental operations learners can perform with regard to a lexical item. Learners can engage in such elaborative processing spontaneously, but they may also benefit from teacher-initiated interventions that stimulate it. Such an intervention may simply cater for the kind of elaboration which the learner would otherwise engage in spontaneously. Additionally, the intervention may extend the scope of the given elaboration strategy to embrace more L2 words, or it might even reveal a pathway for elaborative processing which the learner has not yet considered herself. Let's illustrate this complementary relationship between learner-

initiated and teacher-prompted engagement with reference to the well-known keyword technique.

In this technique, learners associate an L2 word with a familiar word (usually an L1 word), and they use the latter as a key to retrieve the L2 word from memory (see Nation 2001: 311-314, for a review). A suitable keyword shows some phonological (and/or graphemic) resemblance to the target word (otherwise it would not assist in retrieving the L2 word form). The learner then creates a semantic association between the keyword and the target word. In its simplest form, the L2 word could simply be a cognate (in which case the semantic association is very straightforward). Learners are known to resort to cognates spontaneously (e.g. Hall 2002), but a teacher can guide this type of engagement by pointing up suitable cognates for the learners. Crucially, the keyword technique extends beyond straightforward cognates that learners are likely to recognise themselves. Most often – especially when it is teacher-initiated – the technique rests on creative imagery to establish connections between a target word and a keyword whose semantic relatedness is less than direct. For example, a Dutch-speaking learner of English may be advised to associate the target noun *puncture* with the Dutch noun 'puntje' (meaning sharp end) because a punctured tyre and a sharp pointed object can easily be pictured together in a single scene. She may be advised to associate the target verb *frolic* with the Dutch adjective 'vrolijk' (meaning merry) through an image of merrily frolicking chimpanzees. She may be advised to associate the target adjective *brave* with the exclamation 'bravo!', since brave acts merit an applause. And so on. These are the kinds of potentially useful mnemonic associations which the learner might not have turned to when left to her own devices, and so it is the intervention by the teacher (or materials writer) which helps her substantially stretch the use of cognates in the narrow sense.[1]

In the study we report below we shall explore the merits of an intervention which, unlike the keyword technique, is not meant to teach learners a new mnemonic strategy in its own right, but rather to stimulate a type of engagement which learners may sometimes resort to spontaneously but whose full potential is not yet realized. The intervention, in which learners are prompted to consider the degree to which the form of a word might fit the word's meaning, is placed in the framework of Applied Cognitive Linguistics insofar as it relies on the thesis that language is far less arbitrary than has often been assumed.

The contribution of Applied Cognitive Linguistics so far

The principal contribution of pedagogy-oriented Cognitive Linguistics (CL) to L2 vocabulary instruction has been to propose and validate interventions that exploit non-arbitrary dimensions of lexis. More particularly, several authors with a CL background have in recent years suggested methods of showing learners how seemingly distinct meanings and uses of a single word are actually related in ways that are 'motivated' (i.e. explainable in retrospect) (see Boers and Lindstromberg, 2008a, for a review). In the CL view, polysemy is not at all the result of a word taking on new meanings in a random fashion, but rather the outcome of semi-systematic meaning extensions from a word's 'core' sense. For example, familiarity with the physical-motion use of *stumble* may help learners appreciate the 'accidental' nature of a discovery described figuratively as "she stumbled on a piece of evidence" (e.g. Lindstromberg and Boers 2005; Verspoor and Lowie 2003). CL case studies of motivated polysemy and their potential pedagogical applications have not only featured 'content' words, but also such high-frequency 'function' words as prepositions. For example, a chain of meaning extensions can readily be recognised from the literal use of *over* as in "the ball went over the hedge" via a semi-figurative use as in "let's discuss this over a beer" to the fully figurative use as in "they had a dispute over the use of Boolean networks" (cf. Boers and Demecheleer 1998; Cho 2010; Tyler and Evans 2004; Lindstromberg, in press).

In addition to single words, multi-word items such as phrasal verbs and idioms have also been found highly suitable targets for CL approaches to motivated polysemy. For example, the conventionalised, figurative uses of expressions such as *find out (the truth)* and *behind the scenes* can be explained with reference to conceptual metaphors such as KNOWING IS SEEING (e.g. Beréndi, Csábi and Kôvecses 2008; Boers 2000; Boers, Demecheleer and Eyckmans 2004; Condon 2008a; Hu and Fong 2010; Kövecses and Szabó 1996; Skoufaki 2008). The results of several small-scale experiments collectively lend support to CL-inspired vocabulary instruction (see Boers and Lindstromberg 2009, Ch. 5, for a review), although some questions remain as to the generalisability of the findings[2]. CL proponents will need to concede, however, that these proposed approaches focus on *meanings*, and more particularly *extensions* of meanings; it does not help learners much in the way of mapping the initial *form-meaning connection* of the words at the root of those meaning extensions.

At the other extreme, some recent CL-connected endeavours have focused on certain non-arbitrary *formal* properties of lexis. One example of a minimally intrusive intervention in that regard is to briefly alert students to sound patterns such as alliteration in the multiword units

they come across in reading or listening texts (e.g. *bunk bed, to wage war, time will tell*) (Lindstromberg and Boers 2008a). Despite the brevity of the intervention, it appears sufficient to facilitate students' recollection of the lexical makeup of multiword units that happen to display such phonemic repetition. (On similar evidence regarding assonance, see Lindstromberg and Boers 2008b; on the high incidence of phonemic repetition in multiword lexis, see, e.g., Boers and Lindstromberg 2009, Ch. 6).

After having found ways of exploiting motivated meaning-meaning connections (as in polysemy) and motivated form-form connections (as in alliterative multiword units), CL now faces the challenge of finding *form-meaning* motivations that could be used to foster the initial stage of learning (the basic meaning of) new L2 words. Form-meaning matches are commonly referred to as cases of *iconicity.* As one starts to consider the possibility that words might be iconic, one inevitably enters the field of sound symbolism.

Sound symbolism

Sound symbolism refers to non-arbitrary correspondences between phonology and semantics. Some theorists have in fact speculated that early human languages were very much sound-symbolic, and that arbitrary correspondences between semantics and phonology only developed with the need to expand the linguistic means to express more complex messages (e.g. Lecron Foster 1978). Although sound symbolism now appears to be the exception rather than the rule in Indo-European languages (but see below), in many other languages of the world word classes (e.g. so-called ideophones; Doke 1935: 118) remain that display strong systematic sound-meaning mappings (see Nuckols 1999 for a review). Interestingly for language pedagogues, experimental research provides ample evidence that many of the phonological patterns which descriptive linguists have intuited to be iconic are indeed experienced as such by language users generally (see below).

Sound symbolism that is motivated extra-linguistically stands the best chance of being shared by many languages, and may therefore provide useful scaffolding in second language vocabulary learning. Examples of sound symbolism that spring to mind most readily, of course, are onomatopoeia. These indeed show a fair degree of resemblance across languages, although there is obviously also some variation in the way they have been conventionalised (e.g. English *cock-a-doodle-do* vs. Dutch *kukeleku*; English *hiss* vs. Dutch *sissen*). However, several other cases of potentially universal sound symbolism have been examined. One is that

of ‘shape sound symbolism’ (e.g. Ramachandran and Hubbard 2001), for instance, where people are more likely to associate words containing rounded vowels (e.g. “bouba”) with rounded shapes and words containing unrounded vowels (e.g. “kiki”) with angular shapes. The fact that this strong association has been attested not only in adults but also in toddlers (Maurer, Pathman and Mondloch 2006) lends support to the thesis that it is more than a mere by-product of language learning, i.e. of becoming accustomed to hearing and using words which happen to reflect this particular association.

An example of sound-symbolic use of consonants which is language-specific is the phenomenon of so-called phonesthemes (Bloomfield 1933). These are consonant clusters that recur in several words that convey a similar idea. For example, the occurrence of /sw-/ in as many verbs as *swab, sway, sweep, swing, swipe, swirl, swish, swivel,* and *swoop*, is unlikely to be a coincidence. Bloomfield (1933: 245) lists several such sets of words that share a consonant-cluster onset that is potentially sound-symbolic (e.g. *slime, slip, slide* and *crash, crack, crunch*). In a series of experiments, Parault and Schwanenflugel (2006) presented adult speakers of English with sets of unfamiliar words which – on the basis of lists such as Bloomfield’s – were deemed to be either sound-symbolic or non-sound-symbolic. The participants were encouraged to make guesses at the meaning of the words and to match the words with the appropriate definition in a multiple-choice task. The participants obtained significantly better scores on the sound-symbolic words than on the matched controls.

In many languages semantic and/or grammatical classes are signalled pretty systematically through sound. In Japanese, for example, different kinds of motion events are described by means of words that share the same combination of consonants. Imai et al. (2008) report a series of experiments in which participants were asked to match novel instances of Japanese motion words with actions displayed on video. This task posed no problems whatsoever for adult speakers of Japanese, arguably because they were familiar with the language-specific conventions for marking different kinds of motion events. Young Japanese children obtained lower scores, but these were still significantly above chance level. Interestingly, however, a group of adult speakers of English *without* any prior knowledge of Japanese also obtained significantly above-chance scores on the matching tasks. This can only be explained with reference to a sound-symbolic effect that carries over across the two languages. Findings like these give reason to believe that learners’ appreciation of non-arbitrary sound-meaning connections might not only provide scaffolding for vocabulary acquisition in L1 but possibly also for vocabulary learning in L2.

The main research objective of this study is to explore the possibility of turning students' subjective appreciation of the form-meaning motivation of words into a pathway for their engagement with those words. To be clear, although we have resorted to research in sound symbolism as part of establishing the rationale for study, the study itself is not intended to provide objective evidence in support of iconicity *per se*. The nature of said appreciation may not be due to sound symbolism in any strict sense, but also by 'coincidental' resemblances with known words in the L1, the L2 or an additional familiar language (i.e. interlinguistic and/or intralinguistic associations). Hall (2002) reports evidence of learners' spontaneous inclination to relate new L2 words to prior knowledge of word forms - in whichever language they have at their disposal - and their associated meanings. Hulstijn (2001) writes: "If a new word appears to the learner as having a form unrelated to its meaning, it will need more attention and mental elaboration than if it has a transparent appearance." (262) If it is true that language learners implicitly gauge whether a particular word meaning more or less 'fits' the phonological (or graphemic) form they have somehow come to associate with (aspects of) that meaning, then we ask whether this implicit evaluation could be turned into a conscious elaboration strategy, benefiting the initial mapping of meaning onto the form of a new L2 word. This type of elaboration would be explicit, yet minimal in terms of time and cognitive investment.

Research questions

Our general objective led to the following research questions:

1. Does a prompted elaboration of the form-meaning connection of a new L2 word lead to higher learning and retention gains as measured on a *form recall* test in relation to comparison treatments?

2. Does a prompted elaboration of the form-meaning connection of a new L2 word lead to higher learning and retention gains as measured on a *meaning recall* test in relation to comparison treatments?

3. Is the effect of this prompted form-meaning elaboration of a new L2 word influenced by the degree of fit between the form and the meaning as perceived by the learners?

Most studies investigating either L1 word memory (e.g. Eagle and Leiter 1964; Hyde and Jenkins 1969) or L2 word learning (e.g. Barcroft 2002) have instructed subjects to engage with either the meaning of the target words (i.e. semantic elaboration), or the form (i.e. structural elaboration), but rarely to engage explicitly with the form-meaning connection.[4] Expectations concerning the effect of drawing learners' attention simultaneously to meaning *and* form depend on the processing model one adheres to. The Levels of Processing theory (LOP) (Craik & Lockhart 1972) holds that semantic elaboration induces superior processing to structural elaboration, in turn leading to better recall on a test. Transfer-Appropriate-Processing theory (TAP) (Morris, Bransford and Franks 1977), on the other hand, proposes that tasks which stimulate structural elaboration will result in better recollection of form, whereas tasks that encourage semantic elaboration will result in better recollection of meaning. Lastly, Barcroft's (2002, 2003) Type of Processing Resource Allocation (TOPRA) model holds that learners' processing capacity is limited, so attention to one may be at the expense of the other. In this light, as it seems to require dispersion of students' attentional resources, the task of evaluating the form-meaning match of a novel L2 word may arguably be overambitious. However, when participants are encouraged to process items for 'mapping', this could stand them in good stead on a *cued* recall test, where the mapping of meaning onto form or vice versa is exactly what is required. It is this initial, but crucial, step in the word learning process that is under investigation in our study.

DESIGN

General design

The experiment was designed as a between-subjects study which examined the effectiveness of three different learning treatments on the cued recall of 24 new L2 words. The three treatments corresponded to the following three tasks:

1. 'Familiarity assessment'. The participants were told that they were helping the researchers with a norming study to find out which of a series of words were likely to be unknown to learners of a similar level of proficiency. This task was intended not to stimulate any particular elaborative processing of the words.

2. 'Form-meaning-fit assessment'. The participants were asked to evaluate the degree to which they felt the form of L2 words they were presented with matched their meaning. This is the experimental treatment we are putting to the test in this study.
3. 'Utility assessment'. The participants were told they were helping the researchers select words they felt to be sufficiently useful to be included in a new advanced learners' dictionary. This task was intended to stimulate engagement with the meaning of the words, i.e. semantic elaboration.

Like several previous L2 word learning studies (e.g. De Groot and Keijzer 2000; Lotto and de Groot 1998; Schneider, Healy and Bourne 2002), we presented the new L2 words paired with their L1 translations.

The 24 target words were presented to the participants in two sets, separated by a break. After having tackled the first set according to one of the three tasks, the students were given an unannounced post-test. First the L1 word was given and participants were requested to produce the corresponding L2 word (*form* recall, traditionally known as "productive" testing), and subsequently the L2 word was given and participants were asked to provide the corresponding L1 word (*meaning* recall, traditionally known as "receptive" testing).

The second set of words was tackled by the participants along the same task instruction they were given for the first, but this time the participants knew there was going to be a test afterwards. On the one hand, we wanted to include a sufficient number of L2 words to enhance the validity of any conclusions drawn from the results, and to enable us to estimate the scope of applicability of the experimental treatment if it were shown to be effective. On the other hand, springing an unannounced test on the participants after they had been presented with a considerable number of unknown words in a row risked yielding poor scores and thus a floor effect. The result of the set-up is thus a sequential combination of incidental and intentional learning conditions.[4] Let it be clear from the start, however, that the experiment was not designed to compare gains under incidental and intentional learning conditions *per se*. For one thing, the target words in each set were different and not deliberately matched. For another, both sets of words were tackled in one session, and so fatigue was more likely to affect performance on the second set (i.e. the intentionally studied words). It is the effect of the form-meaning-fit-assessment task in comparison with other treatments that we are interested in here. While it may be interesting to find out if this effect – if any – becomes more noticeable if students perform the task purposefully as part of a mnemonic strategy, the results obtained from our study can only be suggestive in that regard.

Two weeks later, a delayed post-test was administered. This consisted of the same test items as the immediate post-test. For a general outline of the experimental design, see Table 1.

Participants

Our participants were 56 university students (20 male, 36 female) in Brussels, Belgium, enrolled in the 1st or 2nd year of an English Language degree. Their average age was 20.2 (SD: 4.7), and their English proficiency was judged by their teachers to approximate level B2 according to the descriptors in the Common European Framework of Reference. Participants' native language was Dutch or a combination of Dutch with another language, mostly French. All participants were highly experienced language learners; in combination with English their university degree involved one other foreign language, and many of them reported working knowledge of at least one more additional language. The experiment was carried out in four intact classes, and students were randomly assigned to one of three treatments: 'familiarity assessment' (n=17); 'form-meaning-fit assessment' (n=20), designed to induce an elaboration of the form-meaning connection; and 'utility assessment' (n=19), devised to encourage semantic elaboration. Two weeks later, 39 participants took part in the delayed post-test.

Given our choice of target words (see below), prior knowledge of these was extremely unlikely. As far as their general proficiency in English was concerned, the three treatment groups were found well matched on the basis of the students' end of semester grades (p=.68).

Target words

As it was essential to control for prior word knowledge, 24 very rare, mostly obsolete words were culled from an obscure word list on the internet (www.*obscurewords.com*). These words were of various lengths, and they belonged to three word classes: nouns, verbs, and adjectives. The 8 nouns used in the study were: *foppotee* 'idiot', *seraglio* 'palace', *welkin* 'sky', *yawd* 'mare', *meed* 'reward', *bandobast* 'settlement', *cant* 'hypocrisy', *mattoid* 'madman'. The 8 verbs were: *blandish* 'flatter', *hie* 'leave', *sough* 'sigh', *tope* 'drink', *madefy* 'moisten', *vitiate* 'damage', *fub* 'postpone', *gledge* 'squint'. Finally, the 8 adjectives included in the study were: *cinnabar* 'dark red', *voluble* 'talkative', *harageous* 'brutal', *gibbous* 'round', *sere* 'dry', *mim* 'modest', *mellifluous* 'harmonious', *luculent* 'clear'.

These stimuli were not manipulated in any way or controlled for the effects of frequency, concept familiarity, imageability, cognate status, phonotactic regularity, or any other psycholinguistic aspect that could inform word memorability. We only ensured the 24 words were grouped into two groups of equal length, i.e. two series of 12 words each - constituting our separate sets which would fuel either the incidental or the intentional learning condition - , comprising a total of 24 syllables each.

The words targeted in our experiment were not selected on the basis of previously identified sound-symbolic patterns either. Neither did any of the authors – who shared the participants' L1 – anticipate that any of the words was likely to call up an L1 cognate. Using a mixed bag of items adds to the ecological validity of the study, and it can help to estimate what fraction (if any) of (English) words are amenable to the type of processing we are putting to the test here. Such an estimate would not be possible if we were to use a pre-selection of likely sound-symbolic items (as was done by, e.g., Parault and Schwanenflugel 2006).

Procedure

The experiment was carried out in the participants' regular computer classroom, during scheduled class hours. At the start of the lesson students were informed, in Dutch, that they were taking part in a survey, and that their answers would have no negative influence on their marks for English. They were invited to sit down at a computer, to fill in a language background questionnaire, and to log onto the software programme (Question Mark Perception, version 3) using the individual log-in and password provided on their questionnaires. Unbeknownst to the students, this automatically – and randomly - assigned each of them to one of the three treatments/tasks: familiarity assessment, form-meaning-fit assessment or utility assessment.

Each target word was then shown in the middle of the screen, accompanied by its translation. The students also heard the target word pronounced once. They were then given 12 seconds to perform the task, which involved making an indication on a five-point Likert scale. Students assigned to the familiarity assessment treatment were asked to indicate the extent to which they were familiar with the words. Students assigned to the form-meaning-fit assessment treatment indicated the degree to which they felt the form/sound of each word

went together with its meaning. Students assigned to the utility assessment treatment rated the words for how useful they felt them to be.

After giving students the opportunity to ask questions about the procedure, a first set of 20 words, including 12 of our target words, was presented to them, and the students individually tackled their respective rating tasks. The first 3 words given were 'dummy' items, so that participants could familiarize themselves with the procedure, as well as to control for primacy effects. To counteract recency effects, the last 2 stimuli given were also excluded from the analyses. For affective reasons, we included in the set three filler words ('pity', 'dwell', 'brave') which we expected the students to be familiar with.

The presentation and rating of the first 20 words was followed by an unannounced, immediate post-test. This testing phase consisted of 2 blocks: a form recall test followed by a meaning recall test. In the form recall test, learners were given 15 questions, each presenting the Dutch translation of an English word they had just evaluated, as well as the first letter of said English word, and asked to produce it. After this test block, learners were presented with 15 questions again, each providing the English word, and learners were then asked to provide the Dutch translation, for which they were not given the first letter. This was the meaning recall test. Both testing blocks were self-paced.

After completing this part of the assignment, participants were asked to note down on their questionnaires whether they had expected to be tested on the vocabulary presented. None of the students reported in the affirmative.

A short break was given, after which participants were invited to proceed with the second series of words. They were again instructed on the screen to carry out the rating exercise as in the previous series but also - since they knew it was to be followed by a post-test - to do their best to learn the words as well as they could. This time, the filler words that we expected the participants to be familiar with were 'herd', 'toss', 'slender'. After the retention test on the second set of words (which followed the same pattern as the test administered after the first set), the students were thanked for their participation and told that they would be informed about their test results in due course.

Two weeks later, 39 original participants were given an unannounced delayed post-test, identical to the immediate one. The entire procedure lasted approximately 70 minutes (learning phase + immediate post-test: 50 minutes; delayed post-test: 20 minutes).

Scoring and analysis

For the meaning recall test, dichotomous scoring was applied: one point was awarded to each word **correctly** and **completely** translated into Dutch, zero points to incorrect translations. In the case of the form recall test, however, we applied a less stringent scoring protocol. As participants were only exposed to each lexical item once, and new words are learnt incrementally, we needed a recall measure that was sensitive to both complete and partial word learning. This measure was supplied by the well-established *Lexical Production Scoring Protocol* (Barcroft 2000), which awards **.25, .50, .75, or 1 point** to each word **partially** or **completely** produced (see Appendix). By giving the first letter of the target words and by asking the participants to try and reproduce the words they had been presented with in the previous stage of the experiment, we prevented the participants from filling in known synonyms of the targets.

We submitted the scores for each testing block to an analysis of variance (ANOVA). To provide an answer to all three research questions we needed to include both an analysis by participants and one by items. As our selection of target words had not been controlled for item effects, we could not simply compare students' test scores obtained under the incidental with those under the intentional learning condition (given the possibility that item effect might interact with learning condition). Therefore, the analysis by participants was run on the test scores for the two sets of 12 words separately, with treatment ('familiarity assessment', 'form-meaning-fit-assessment, 'utility assessment') as between-subject factor and retention interval (immediate, delayed) as within-subject factor. The ANOVA by items added learning condition (incidental, intentional) as a between-subject factor to this model, but since the items were not crossed across learning condition, items were *nested* within that factor to control for potential item effects.[5]

Results

Analysis by participants

Form Recall Scores

In the incidental learning condition (first series of words), the analysis reveals a significant effect of treatment, with a moderate effect size: $F(2,89)=3.254$, $p < .05$, $\eta 2 = .068$. Table 2

shows recall rates across conditions. It reveals that the form-meaning-fit-assessment group outperforms the two other groups in both post-tests. The delayed post-test amplifies the divergence between our experimental group and the utility-assessment group.

The average recall rates for words in the intentional learning condition (i.e. our second series of words) across treatments and post-tests are shown in Table 3. The form-meaning-fit-assessment group outperforms the other two groups in both the immediate and the delayed post-test under this condition as well. The divergence falls short of statistical significance, however, which could be due to the high standard deviations recorded.

Meaning Recall Scores

There are no significant main effects in the incidental learning condition. Meaning recall scores are higher than form recall scores generally, yet remarkably similar across all treatment groups: participants in the familiarity assessment group score an average of 7.34 (1.98) across both post-tests, those in the form-meaning-fit group achieve a mean of 7.24 (1.58), and those in the utility assessment group an average of 7.06 (1.93). Moreover, the word meanings are retained remarkably well between the immediate and the delayed post test; there is **no** effect of retention interval (p = .503).

As regards our second series of 12 words (intentional learning condition), the analysis does reveal a significant effect of retention interval, $F(1,87)=11.367$, $p = .001$, $\eta^2 = .116$. On the immediate post-test, the form-meaning-fit-assessment group returns a higher mean score (5.61) than the other two groups: 4.59 (familiarity assessment) and 5.21 (utility assessment). However, in the delayed post-test, the utility-assessment group shows less attrition than the others: 3.7 vs. 3.23 (form-meaning-fit-assessment) vs. 3.23 (familiarity assessment). None of these differences are statistically significant, however, nor do they yield a significant interaction between treatment and retention interval.

Analysis by items

Form Recall Scores

The item analysis confirms the significant effect of treatment: F(2, 44)=6.654, p =.003, η2=.232. Post-hoc paired comparisons (Tukey) indicate that the difference between the form-meaning-fit-assessment group and the familiarity-assessment group is significant at p =.022, and with the utility-assessment group even at $p < .0001$. Table 4 presents the mean form recall scores by items across all conditions.

The effect is not equally strong for all items, however. Although item effect falls (just) short of statistical significance (p = .066), it seems likely that the effect of the task(s) is likely to be qualified by the properties of individual items.

Our treatment variable is borderline significant in interaction with learning condition (i.e. incidental versus intentional) (p =.059). This is probably due to the divergence in recall rates obtained under the familiarity-assessment treatment: they are relatively high for the incidentally learned words (i.e but comparatively low for the intentionally learned words. Table 4 reveals a similar interaction between treatment and condition in the *delayed* form recall scores: it is with regard to the 'intentionally' learned words that the form-meaning-fit-assessment task appears mnemonically more effective in the long term than both of the other tasks. What is surprising is that words were retained less well through the utility-assessment task than the familiarity-assessment task, while we expected the former to stimulate more engagement with the words.

Meaning recall scores

The analysis by items uncovered no main effect of treatment here. As expected, word meanings were more likely to be recalled immediately after the presentation of the words than two weeks later, but the attrition rate (which is similar in the three groups) is much less pronounced than in the case of form recall. It may look surprising that the word meanings of the 12 words presented in the incidental learning condition were much better remembered than the 12 words shown in the intentional learning condition, with means of 7.19 and 4.25, respectively (F(1,22)=20.662, p <.0001, η2 =.484). However, as was mentioned above, the lack of uptake despite the anticipation of a post test may in this experiment simply be due to

fatigue on the part of the students towards the end of the session, and we cannot be sure that the learning burden of the two sets of target words was equivalent in the first place. Table 5 shows that the retention interval had a much larger effect on the meaning recall of items that were learnt intentionally than on the items in the incidental condition; performance on the incidentally learned items nudged downwards ever so slightly, whereas performance on the intentionally learned items plummeted.

Recall rates for 'motivated' items

The 20 participants in the form-meaning-fit-assessment group rated each word according to how well they felt its meaning to fit its form. The average Likert score (on a scale from 1-5) given to our 24 stimuli was 2.71 (1.35). We used this mean as a cut-off point to organize our items into a 'more motivated' and a 'less motivated' group.

The 12 words in our 'more motivated' group were (in descending order according to motivation ratings, standard deviations between brackets): *harageous* 'brutal': 4.2 (.83), *voluble* 'talkative': 3.7 (1.17), *luculent* 'clear': 3.58 (1.31), *sere* 'dry': 3.47 (1.22), *mattoid* 'madman': 3.37 (1.21), *sough* 'sigh': 3.35 (1.23), *foppotee* 'idiot': 3.25 (1.25), *welkin* 'sky': 3.2 (.894), *seraglio* 'palace': 3.2 (1.32), *mellifluous* 'harmonious': 3.11 (1.37), *mim* 'modest': 2.79 (1.48), *blandish* 'flatter': 2.74 (1.37).

The remaining 12 words in our 'less motivated' group were: *tope* 'drink': 2.5 (1.43), *gledge* 'squint': 2.42 (.961), *madefy* 'moisten': 2.42 (.961), *gibbous* 'round': 2.3 (1.13), *hie* 'leave': 2.2 (1.47), *fub* 'postpone': 1.95 (1.05), *meed* 'reward': 1.95 (1.18), *cinnabar* 'dark red': 1.85 (.745), *vitiate* 'damage': 1.84 (1.07), *bandobast* 'settlement': 1.84 (1.12), *cant* 'hypocrisy': 1.84 (1.35), *yawd* 'mare': 1.8 (1.06).

Coincidentally, the 12 words from each of our 2 learning conditions (incidental and intentional) are equally distributed across these two groups, suggesting the items in each condition would have appeared comparably motivated to our raters.

We analysed the item recall scores using analysis of variance (ANOVA), this time including motivation strength (2) as between-subjects factor, and treatment (3) and retention interval (2) as within-subjects factors. Table 6 shows the recall scores for the 'less motivated' and 'more motivated' items per treatment, collapsed across retention intervals. It reveals that most recall scores are remarkably similar across the board, especially as regards our 'less motivated' items. This is confirmed by our analyses, which reveal neither a main effect of strength of motivation, nor a significant interaction effect between treatment and motivation

strength. However, a divergence between *form* recall scores is found among the 'more motivated' items: form recall scores for those items in the form-meaning-fit-assessment group are higher (4.54) than in the familiarity-assessment group (3.69), and considerably higher than in the utility-assessment group (3.09).

When we split our data file and take an exclusive look at the form recall scores for the 'motivated' items, we do obtain a highly significant effect of treatment with a large effect size, $F(2,66)=8.254$, $p = .001$, $\eta 2 = .200$. Graph 1 shows the effect of treatment on the form recall rates of 'more motivated' items compared to the 'less motivated' items. Post-hoc analyses (Tukey) reveal that it is the divergence between the form-meaning-fit-assessment and utility-assessment groups that is most significant ($p < .0001$). This is not just a congruence effect; if we run the same separate analysis on the items rated comparatively 'useful' by our participants in the latter group, no such treatment effect is revealed for either form or meaning recall rates.

Discussion

This study aimed to compare the short-term and medium-term effectiveness of three different treatments for the receptive and productive recall of unknown L2 words. Each of the treatments involved a rating exercise pertaining to the words, which for the participants in the familiarity-assessment group was intended to induce little elaboration of any particular kind, in the form-meaning-fit-assessment group an engagement with the form-meaning connection, and in the utility-assessment group a semantic elaboration. The mnemonic effect of each treatment task was measured using immediate and delayed form and meaning recall tests over a two-week interval. The main findings will be discussed with reference to the 3 research questions.

1. Does the prompted elaboration of the form-meaning connection of a new L2 word lead to higher learning and retention gains as measured on a *form recall* test in relation to comparison treatments?

Our form recall scores reveal a significant effect of treatment. The form-meaning-fit-assessment task yields the highest form recall scores, in the immediate as well as the delayed post-test. This can probably be attributed to the engagement with word form that this task is

likely to induce in participants, in line with the principles of Transfer Appropriate Processing (TAP) (Morris, Bransford and Franks 1977). Cued form recall on a test is the first step on the way to fluent, productive word knowledge, and as such it charts the most challenging and elusive aspect of word learning (Laufer 2005, 2006; Schmitt 2008). If the pedagogical aim is to enable learners to eventually use words productively, engagement with words must include a structural component. Moreover, an exclusive focus on semantic elaboration is likely to inhibit word form learning (Barcroft 2002, 2003), and this is corroborated by the comparatively poor scores on the form-recall tests generated by the utility-assessment task. In fact, even the familiarity-assessment task, which was meant not to stimulate any particular elaboration, seems to yield better form recall than the utility-assessment task (although the difference in scores is not statistically significant). This substantiates the thesis that semantic elaboration does not foster word *form* learning.

The superiority of our experimental treatment held out over time, although there was no interaction effect between treatment and retention interval, indicating attrition rates were similar across groups.

2. Does the elaboration of the form-meaning connection of a new L2 word lead to higher learning and retention gains as measured on a *meaning recall* test in relation to comparison treatments?

By rating words in terms of their usefulness to learners of English, participants in one of the two control conditions, namely the utility-assessment group, were encouraged to think about the meanings of these items, and possibly of available synonyms, thus inducing semantic elaboration. According to LOP-theory (Craik and Lockhart 1972), this engagement with meaning is expected to lead to superior results in comparison to engagement with form. TAP-theory (Morris, Bransford and Franks 1977) posits that test scores are enhanced when testing measure is congruent with learning condition. Hence, both theories predict comparatively good scores on the meaning recall test after the utility-assessment task. On the other hand, learners who performed the form-meaning-fit-assessment task were coaxed into mapping meaning onto form, and so a cued meaning recall test must be congruent with that learning mode also. The meaning recall scores turned out uncannily similar across the groups. This suggests that our experimental treatment does certainly not impede retention of semantics as compared to other treatments.

The performance of the familiarity-assessment group is perhaps the most surprising of all, especially in terms of the test scores on the items tackled in the incidental learning condition. This does not necessarily challenge the view that word meaning learning is enhanced by engagement, as we cannot be sure that the participants in the familiarity-assessment task did not engage in any elaborative processing, despite the nature of the task.

Attrition rates were similar across the board, as there was no interaction between treatment and retention interval.

Before concluding our discussion with regard to form and meaning recall scores, we need to address how our findings relate to the TOPRA model (Barcroft 2002, 2003) of lexical processing. After all, the model predicts that attending to both form and meaning of lexical items leads to the dispersion of students' attentional resources. This would imply that meaning recall scores are impaired by treatments that include attention to form as compared to treatments that target meaning exclusively. Yet this is not corroborated by our meaning recall results. Likewise, it may suggest that treatments that include attention to meaning would inhibit form recall scores as compared to treatments that solely fixate on form. As our study does not include an exclusively 'form-focused' elaboration treatment, this cannot be substantiated at present. But uur results do not actually contradict the TOPRA-model either. Firstly, our experimental treatment does not induce participants to attend to form and meaning separately, but instead it encourages them to map them together - 'processing for mapping' in the model. Secondly, our dependent measures were *cued* recall tests. These necessitate the mapping of meaning onto form, in contrast to *free* recall tests, which, it could be argued, do not chart L2 word learning at all.[6] Of all the treatments induced, it is this 'processing for mapping' that is actually most congruent with a cued form and meaning recall tests, and its success is therefore not so unexpected after all. Our results do show that the high form recall scores obtained by the form-meaning-fit-assessment group are not at the expense of meaning recall scores; unlike the utility-assessment task, our experimental task does not seem to trade off form recall for meaning recall. This suggests that our proposed type of elaboration is effective in enhancing both form and meaning recall, at least for our group of learners.

3. Is the effect of this prompted form-meaning elaboration of a new L2 word influenced by the degree of fit between the form and the meaning as perceived by the learners?

In terms of form recall scores, the form-meaning-fit-assessment treatment was more effective for items rated linguistically motivated by the participants than items that were less so (4.54

vs. 3.73). This does not seem to be due to giving positive ratings to these words *per se.* If this were the case, the words rated as useful by participants in the utility-assessment group should then have yielded better scores that those rated less useful, and not such differences were observed in the test scores of these participants.

Rather, the question arises as to whether the success of these motivated items in the experimental treatment points to a possible congruence between internal and external salience. As proposed by Sharwood-Smith (1991, 1993), linguistic features that are noticed autonomously by the learner have internally-created salience, necessitating no further instructional intervention. Externally-created salience, on the other hand, is generated through pedagogical intervention, and is required when surface features are not noticed autonomously by the learner, whose attention then needs to be drawn to said features. More empirical research is needed to ascertain whether spontaneously occurring salience has a different learning effect than when salience is deliberately engineered through instruction (Sharwood-Smith 1991:121). However, it would certainly make sense to consider instruction optimal when it can capitalize on learners' internal salience, forging a kind of harmony between the two types of salience (cf. Park & Han 2007:110). (See also our brief discussion of the keyword technique in the introduction to this article.) If the form-meaning connection of some words appears more 'naturally' motivated to learners than others, explicitly attending to this apparent motivation by way of a minimal, pedagogical intervention would certainly seem to capitalize on its internal salience. Moreover, even though word motivation might be an idiosyncratic affair at times, the standard deviations recorded for our Likert motivation ratings are moderate, pointing to an acceptable inter-rater agreement for many of these items. This suggests that teachers' subjective appreciations of form-meaning motivations could by and large be relied upon and provide the basis for short pedagogical interventions. More research needs to be done to ascertain this.

Incidentally, and interestingly in this respect, our filler items, i.e. *toss, dwell, herd, pity, brave,* and *slender*, scored very highly on this Likert scale for linguistic motivation; median 5. It suggests that language users collapse 'form' and 'meaning' entirely when a word is known, i.e. once its form-meaning connection is deeply entrenched and fluent.

Limitations

A number of possible restrictions constrain the generality of our study's findings. First of all, the groups of participants were relatively small, and each participant received only one of the three types of treatments. We were compelled to use this between-subject design rather than comparing the effect of different treatments on the same individuals, because a pilot study had revealed that each of the tasks encroached too much on the following if assigned in combination. Given the small size of the groups, individual differences in learning styles may accidentally have impacted the relative effectiveness of this or that treatment.

Secondly, whether the rating tasks given to the learners are good operationalizations of the constructs of 'zero elaboration', 'form-meaning elaboration' and 'meaning elaboration' is subject to debate. We deemed it essential for each treatment to involve a similar mental and motor task; i.e. assess the word on a scale, but we wanted the quality and quantity of the induced elaboration to be different. The most important concern in this respect is raised by the familiarity-assessment task, which we thought came close to a 'zero elaboration' treatment. Yet assessing to what extent one 'knows' a word might imply a type of elaboration after all, which, if induced, would involve attending to both meaning and form, as it encourages a reflection on the learner's whether meaning and form are already associated with one other, presumably on the basis of possible previous encounters with the word in question.

Lastly, operational constraints also meant that our 12 words in the incidental learning condition were different from our 12 words in the intentional learning condition, yet identical for each treatment group, and presented in the same order. It may come as a surprise that recall scores were higher in the incidental learning condition than in the intentional learning condition (see the tables in the results section), but the effect of intentional learning is likely to have been eclipsed by other variables, such as primacy effects, fatigue and memory crowding. Given the different variables at play, we cannot say at this stage whether the effectiveness of our proposed pedagogical intervention is influenced by type of motivation for learning.

Conclusion and implications for vocabulary learning and teaching

Our study suggests that an evaluation of the linguistic motivation of a word's form-meaning link can be an effective form of elaboration. As such, it could be turned into a conscious,

simple, and time-efficient word learning strategy, providing an extra pathway to strengthen the form-meaning mapping of a new word. Lexical items whose form-meaning link is perceived to be motivated are particularly amenable to it. What's more, the 'processing for mapping' induced by our form-meaning-fit-assessment appears equally beneficial for the learning and retention of word meaning as word form, with neither aspect of word knowledge being established at the expense of the other.

In terms of vocabulary teaching, the appreciation of the form-meaning motivation of words does not need to be scientific or universally shared; as long as it is there for a particular learner, it can have mnemonic potential (cf. Croft 1978). Still, the degree of agreement among our participants on which of the items were relatively motivated suggest that – for lack of a thesaurus of 'sound-symbolic' words – teachers could rely on their subjective appreciations of form-meaning motivations to justify short pedagogical interventions.[7] We propose that an elaboration of the form-meaning connection can help to enhance the mapping of meaning onto form and vice versa. As there may not be a 1:1 correspondence between meanings in the L1 and the L2, further semantic fine-tuning could still be required, and this initial mapping may not be sufficient to achieve fluent word knowledge, but it does constitute the first essential step – or leap – towards it.

Teachers' interventions to tap this resource can be very simple and brief. For example, on encountering an unfamiliar L2 word during a classroom activity, the teacher might pronounce this word in a tone of voice that is compatible with the word's 'connotations', and with exaggerated articulation or lengthening of certain phonological features to hint at a certain degree of iconicity. One can do this with words as varied as *slime*, *whisper* and *smooth*. This kind of minimal intervention on an as-the-opportunity-arises basis is reminiscent of what Lindstromberg and Boers (2008a) have found to be measurably effective when applied to alliterative phrases. The auditory stimulus for elaboration can of course go hand in hand with others, such as awareness-raising of morphological clues (as *pro* and *long* in *prolonged*) and semi-cognates (as 'vol' in *voluptuous* – Dutch 'vol' means full). Student involvement can easily be invited when a phonestheme is encountered: on encountering the word *swirl* the teacher can easily ask students to pool other sw- verbs they happen to know and contemplate what semantic feature they have in common with the new word.

But also in exercises designed to teach vocabulary more directly, potential form-meaning fits can perhaps be put to good use, and more particularly to help avoid the risk of erroneous initial form-meaning correspondences. For example, matching exercises (e.g. connecting L2 words to corresponding L1 words or corresponding L2 synonyms) can be pedagogically

sound only if the risk of erroneous matches is minimized. One way of doing this is by making sure that only some of the items in the exercise are new to the learners. A supplementary way suggested indirectly by the results of our study may be to reduce blind guessing by selecting words for the exercise that display a certain degree of form-meaning-fit. Whether a given word is suitable in this respect may have to be tested through piloting, however. This piloting would ideally have to be done with same-population students. As we mentioned in the introduction, whether or not the form-meaning connection of a given L2 word is perceived as 'motivated' is likely to be influenced not only by universally shared sound symbolism but also by cognate effects.

What we have attempted to do in this contribution is to complement the existing CL approaches to teaching L2 vocabulary, most of which exploit meaning-meaning motivation (as in polysemy) and some form-form motivation (as in alliterative word partnerships), by exploring the pedagogical potential of form-meaning motivation. Our intention has not been to provide evidence of iconicity *per se*, but rather to investigate whether making learners consider the possibility of such iconicity is a fruitful way of stimulating engagement with L2 words. We believe the preliminary results are encouraging.

Suggestions for further research

A qualitative follow-up study is warranted to answer a number of questions. We wish to know whether the advantage of the experimental treatment for form recall is possibly informed by an *affective* dimension that the other two treatments might not afford. Reactions by the students indicated that they enjoyed the rating exercise, but further study could investigate to what extent they also appreciate the relevance of it, and how justified or natural it seems to them. Furthermore, the validity of our conclusions could be explored for students at lower levels of proficiency and with different learner characteristics. Finally, qualitative data could offer an insight into how different learners 'motivate' their ratings, whether it be on the on the basis of sound symbolism, loanwords and cognates, or idiosyncratic associations.

Acknowledgements

The authors would like to thank Hélène Stengers for her invaluable assistance with QuestionMark Perception (3). We also thank two anonymous reviewers for their suggestions, which have helped us make this a more reader-friendly article – we hope – than the original version was. Last but not least we are grateful to the students who took part in the experiment.

Notes

1. Some investigations of the keyword technique suggest that its efficiency decreases as the semantic link between keyword and target word becomes far-fetched (e.g. Wang and Thomas 1995).
2. Laufer (2005) likens the difference between the systematic teaching of pre-selected vocabulary (e.g. through word translation exercises) and the more occasional targeting of vocabulary during communicative activities to the distinction that in contexts of grammar instruction has become known as 'focus on form*s*' and 'focus on form', respectively (Long 1991). It is not always clear where CL proposals for language pedagogy are to be situated in this framework, although the input materials used in most CL experiments to date seem to suggest leanings toward form***s***-focused instruction.
3. To operationalize the construct of semantic elaboration, a significant number of studies have instructed participants to rate the experimental words in terms of the pleasantness of their *referents* (e.g. Barcroft 2002, in line with previous L1 experiments). Whether participants are able to divorce meaning entirely from the formal properties of these words in this kind of exercise, however, has never been ascertained.
4. In accordance with Hulstijn (2001; 266-267), we have operationalized the difference between incidental and intentional learning as the absence or presence of a warning to participants that they will be tested afterwards.
5. A nested design is an experimental design in which the variables are not crossed, but have an implicit hierarchy. In our case, words 1-12 are nested (i.e. embedded) within the 'incidental' learning condition, and words 13-24 are nested within the 'intentional' learning condition.
6. Recalling the forms of target L2 words freely after a treatment is no evidence of word learning if you do not know what they mean, and recalling the meanings of target L2 words freely (i.e by producing known L1 words) suggests you have a good episodic memory.

7. Given the likelihood that a fair number of subjective form-meaning associations might be due to cognate effects in a broad sense, the chances that a teacher's appreciation of such appreciations will correspond to the learners' may be reduced if the teacher is unfamiliar with the learners' L1.

References

Barcroft, J. 2002. Semantic and structural elaboration in L2 lexical acquisition. *Language Learning, 52*: 323-363.

Barcroft, J. 2003. Effects of questions about word meaning during L2 Spanish lexical learning. *The Modern Language Journal, 87*: 546-561.

Beréndi, Csábi and Kôvecses. 2008. Using conceptual metaphors and metonymies in vocabulary teaching. In Boers, F. & Lindstromberg, S. (eds.). *Cognitive Linguistic Approaches to Teaching Vocabulary and Phraseology*, 65-99. Berlin: Mouton de Gruyter.

Bloomfield, L. 1933. *Language.* New York: Holt, Rinehart and Winston.

Boers, F. 2000. Metaphor awareness and vocabulary retention. *Applied Linguistics* 21: 553–571.

Boers, F., and M. Demecheleer. 1998. A cognitive semantic approach to teaching prepositions. *English Language Teaching Journal* 53: 197–204.

Boers, F., M. Demecheleer, and J. Eyckmans. 2004. Etymological elaboration as a strategy for learning figurative idioms. In *Vocabulary in a Second Language: Selection, Acquisition and Testing*, P. Bogaards and B. Laufer. (eds.), 53–78. Amsterdam/Philadelphia: John Benjamins.

Boers, F., J. Eyckmans, and H. Stengers. 2006. Means of motivating multiword units: Rationale, mnemonic benefits and cognitive-style variables. In *EUROSLA Yearbook 6*, S. Foster-Cohen, M.M. Krajnovic and J.M. Djigunovic (eds.), 169–190. Amsterdam/Philadelphia: John Benjamins.

Boers, F., and S. Lindstromberg. 2009. *Optimizing a Lexical Approach to Instructed Second Language Acquisition*. Basingstoke: Palgrave Macmillan.

Cho, K. 2010. Fostering the Acquisition of English prepositions by Japanese learners with networks and prototypes. In *Fostering Language Teaching Efficiency through Cognitive Linguistics*, S. De Knop, F. Boers, and T. De Rycker (eds.), 259-275. Berlin: Mouton de Gruyter.

Condon, N. 2008a. How cognitive linguistic motivations influence the phrasal verbs. In *Cognitive Linguistic Approaches to Teaching Vocabulary and Phraseology*, F Boers and S. Lindstromberg (eds.), 133-158. Berlin: Mouton de Gruyter.

Condon, N. 2008b. Investigating a cognitive linguistic approach to the learning of English phrasal verbs. Unpublished PhD. Dissertation. Université Catholique de Louvain-la-Neuve.

Craik, F.I.M., & Lockhart, R.S. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour, 11*: 671-684.

Croft, L.B. 1978. The mnemonic use of linguistic iconicity in teaching language and literature, *Slavic and East European Journal* 22: 509-518.

De Groot, A.M.B, & Keijzer, R. 2000. What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning, 50* (1): 1-56.

Doke, C.M. 1935. *Bantu Linguistic Terminoloy.* London: Longman, Green.

Eagle, M., & Leiter, E. 1964. Recall and recognition in intentional and incidental learning. *Journal of Experimental Psychology, 68* (1): 58-63.

Hall, C.J. 2002. The automatic cognate form assumption. Evidence for the parasitic model of vocabulary development. *IRAL 40*, 69-87.

Hu, Y.H., and Y.Y. Fong. 2010. Obstacles to CM-guided L2 idiom interpretation. In *Fostering Language Teaching Efficiency through Cognitive Linguistics*, S. De Knop, F. Boers, and T. De Rycker (eds.), 293-317. Berlin: Mouton de Gruyter.

Hulstijn, J.H. 2001. Intentional and incidental second language vocabulary learning: a reappraisal of elaboration, rehearsal and automaticity. In Robinson, P. (ed.), *Cognition and Second Language Instruction*, 258-286. Cambridge: Cambridge University Press.

Hyde, T.S., & Jenkins, J.J. 1969. Differential effects of incidental tasks on the organization of recall of a list of highly associative words. *Journal of Experimental Psychology, 82* (3): 472-481.

Imai, M., S.Kita, M.Nagumo, H.Okada. 2008. Sound symbolism facilitates early verb learning, *Cognition* 109: 54-65.

Kövecses, Z., and P. Szabó. 1996. Idioms: A view from Cognitive Semantics. *Applied Linguistics* 17: 326–355.

Laufer, B. 2005. Focus on Form in second language vocabulary acquisition. In *EUROSLA Yearbook 5*, S. Foster-Cohen (ed.), 223–250. Amsterdam/Philadelphia: John Benjamins.

Laufer, B. 2006. Comparing Focus on Form and Focus on FormS in second language vocabulary teaching. *The Canadian Modern Language Review, 63* (1), 149-166.

Laufer, B., and J. Hulstijn. 2001. Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics* 22: 1-26.

Lecron Foster, M. 1978. The symbolic structure of primordial language. In *Human evolution biosocial perspectives,* S.L. Washburn and E.R. McCown (eds.), 77-122. Merlo Park, CA: Benjamin/Cummings Publishing.

Lindstromberg, S. In press. *English Prepositions Explained.* 2nd ed. Cambridge: Cambridge University Press.

Lindstromberg, S., and F. Boers. 2005. From movement to metaphor with manner-of-movement verbs. *Applied Linguistics* 26: 241–261.

Lindstromberg, S., and F. Boers. 2008a. The mnemonic effect of noticing alliteration in lexical chunks. *Applied Linguistics* 29: 200–222.

Lindstromberg, S., and F. Boers. 2008b. Phonemic repetition and the learning of lexical chunks: The mnemonic power of assonance. *System* 36: 423-436.

Long, M.H. 1991. Focus on form: A design feature in language teaching methodology. In *Foreign Language Research in Cross-Cultural Perspective,* Kees de Bot, Ralph B. Ginsberg, and Claire. Kramsch (eds.), 39–52. Amsterdam/Philadelphia: John Benjamins.

Lotto, L., & De Groot, A.M.B. 1998. Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning, 48* (1): 31-69.

Maurer, D., T. Pathman, and C.J. Mondloch. 2006. The shape of boubas: sound-shape correspondences in toddlers and adults. *Developmental Science 9*: 316-322.

Morris, C.D., Bransford, J.D., & Franks, J.J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*: 519-533.

Nation, I.S.P. 2001. *Learning Vocabulary in Another Language.* Cambridge: Cambridge University Press.

Nuckols, J.B. 1999. The case for sound symbolism. *Annual Review of Anthropology 28*: 225-252.

Parault, S.J, and P.J. Schwanenflugel. 2006. Sound symbolism: a piece in the puzzle of word learning. *Journal of Psycholinguistics Research 35*: 329-351.

Ramachandran, V.S., and E.M. Hubbard. 2001. Synaesthesia: a window in to perception, thought and language. *Journal of Consciousness Studies 8*: 3-34.

Park, Eun Sung, & Han, Zhaohong. (2007). Learner spontaneous attention in Second

Language input processing: An exploratory study. In Han, Zhaohong, & Park, Eun Sung. *Understanding Second Language Process*, 106-132. Clevedon: Multilingual Matters.

Schmitt, N. (2008), Review article: Instructed second language vocabulary learning, *Language Teaching Research* 12 (3): 329-363.

Schneider, V.I., Healy, A.F., & Bourne, L.E. (2002) What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language, 46*: 419-440.

Sharwood-Smith, M. (1991). Speaking to many minds: on the relevance of different types of language information for the L2 learner. *Second Language Research, 7* (2): 118-132.

Sharwood-Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition, 15*: 165-179.

Skoufaki, S. 2008. Conceptual metaphoric meaning clues in two idiom presentation methods. In *Cognitive Linguistic Approach to Teaching Vocabulary and Phraseology*, F. Boers and S. Lindstromberg (eds.), 101-132. Berlin/New York: Mouton de Gruyter.

Tyler, A, and V. Evans. 2004. Applying cognitive linguistics to pedagogical grammar: The case of *over*. In *Cognitive Linguistics, Second Language Acquisition, and Foreign Language Teaching*, M. Achard and S. Niemeier (eds.), 257–280. Berlin/New York: Mouton de Gruyter

Verspoor, M., and W. Lowie. 2003. Making sense of polysemous words. *Language Learning* 53: 547–586.

Wang, A.Y, and MH.Thomas. 2005. Effect of keywords on long-term retention: help or hindrance? *Journal of Educational Psychology 87* (3), 468-475.

Appendix

Lexical Production Scoring Protocol- Written (LPSPS-written)

0.00 points	0.25 points	0.50 points	0.75 points	1 point
None of word is written; this includes: •nothing is written •the letters present do not meet any "for 0.25" criteria •English word only is written	¼ of word is written; this includes: •any 1 letter is correct •25-49.9% of the letters are present •correct # of syllables	½ of word is written; this includes: •25-49.9% of letters correct •50-74.9% of letters present	¾ of word is written; this includes: •50-99.9% of letters correct •75-100% letters present	Entire word is written; •100% letters correct

"Correct" refers to any letter written and placed in its correct position within a word; "present" refers to any letter written but not placed in its correct position.

Adopted from Barcroft 2002: 263.