sciendo
ExELL (Explorations in English Language and Linguistics)
11.1 (2023): 1-34
DOI: 10.2478/exell-2023-0005
Original scientific article

Acquisition of English nominal suffix *-er* by advanced EFL learners: a view from usage-based perspective

Višnja Pavičić Takač & Gabrijela Buljan

University of Osijek, Osijek, Croatia

Abstract

The present study investigated advanced Croatian EFL learners' knowledge of five meanings of the English nominal (deverbal) suffix *-er*. It probed their ability to comprehend and produce corpus-rare and presumably unentrenched *-er* nouns in their prototypical agent and instrument meanings and their non-prototypical patient, locative, and causative meanings. It was hypothesized that participants would deal effortlessly with agent and instrument meanings of the low-frequency nouns since the corpus-attested high type frequency of *-er* agents and instruments, among others, suggests the existence of productive corresponding schemas. We hypothesized that participants would struggle with patient, locative and causative meanings of the low-frequency nouns since the corpus-attested low type frequency of the three functions arguably does not support their association with *-er*. A recognition and a production test were administered to two separate groups of English majors at a Croatian public university (n = 131). Results confirm general usage-based predictions about better performance with low-frequency agent and instrument *-er* nouns. However, a detailed examination reveals unexpected results, which confirm that frequency, however important, is not the only factor to include in a future model of EFL learners' derivational proficiency.

Key words: usage-based model; type frequency; token frequency; nominal suffix *-er*.

1. Introduction

The English nominal suffix *-er* has received much attention from (applied) linguists. One reason is its semantic diversity (Lieber, 2004; Marchand, 1969; Panther & Thornburg, 2002; Ryder, 1991). It builds nouns which denote various entity types (people, plants, events, etc.) and represent various semantic roles (agents, instruments, locations, etc.), with many *-er* nouns having multiple meanings themselves. Agentive *-er* was also identified as one of the earliest and "strongest" derivational suffixes in early children's lexicons (Clark, 2014: 426), often causing overgeneralization errors (Clark & Cohen, 1984). The reason English-speaking children acquire agent and instrument -*er* easily is because it fits the general principles of morphological acquisition: semantic transparency of components, simplicity of form and productivity, i.e. type frequency of *-er* in ambient language (Clark, 2014). The presence of many different *-er* agents and instruments in the linguistic environment gives children the confidence to consistently map these predictable meaning(s) onto predictable form, a process fundamental to language acquisition in general (Tomasello, 2003; Bybee, 2010).

In this study, we explore whether similar expectations hold of advanced Croatian EFL learners, but focus on a fuller set of meanings: the corpus-frequent agents and instruments, jointly referred to as 'central meanings' (CMs) and the corpus-rare locatives (*diner* 'place to dine'), patients (*loaner* 'something loaned') and causes (*screamer* 'something that makes one scream'), jointly 'non-central meanings' (NCMs).[1] Although first language (L1) and foreign language (FL) acquisition differ in important respects (Ellis, 2002a, 2002b), both involve implicit frequency-sensitive learning (R. Ellis, 1994; Lowie & Verspoor, 2004; Lowie, 2005; Robinson & Ellis, 2008). Therefore, our goal is to establish whether advanced EFL learners' knowledge of CMs and NCMs of *-er* reflects their different objective frequencies. Since the NCMs are also semantically transparent and formally just as simple as agents and instruments, we assume that their low type frequency may delay or even preclude the recognition and use of *-er* in novel locative, patient and cause nouns by advanced EFL learners (Lowie, 2005). By "novel" we do not mean new to the English-speaking community, but probably unfamiliar/unentrenched with EFL learners since they are token-rare, at least according to the corpus studied.

We focus on deverbal *-er* nouns because, unlike denominals, they are maximally transparent. The base verb names the event in which the referent participates as agent, instrument etc. Since we did our best to select test nouns with semantically transparent and familiar verb bases, which are also more token-frequent than the -er nouns themselves (see below and Method), much of what our participants (do not) know about the low-frequency nouns could be attributed to (imperfect) knowledge of the suffix.

We focus on *low-frequency* nouns for several reasons. First, there is less research on frequency effects for low-frequency structures (Divjak & Caldwell-

[1] There is evidence of some NCM productivity, cf. *dunker*$_{\text{PAT}}$ 'cookies', *refresher*$_{\text{LOC}}$ 'bathroom', etc. "Whereas it is generally the case that the derivational suffix *-er* is an agentive and instrumental affix … we can learn from corpus data that its semantics is more complex. Consider, for example, this citation from COCA: (3) *Outdoor Life 2005*: I had taken bears before and had been hunting for several years for a truly outstanding bear, and here one was standing broadside at 20 yards. I didn't have to think twice about this bear. It was a **shooter** [emphasis in original]" (Lieber, 2014: 90).

Harris, 2015). Second, they allow testing EFL learners' ability to generalize beyond well-known examples. Token-frequent nouns may reveal little since they may be well-entrenched and item-familiar (Bybee, 2007; Nagy et al., 1993). In fact, it has been argued that the relative frequency of base vs. derivative is more relevant for the perception of morphological structure and for the productivity of WF patterns than the overall token frequency of the complex words (Hay, 2001). Regardless of their overall token frequency, complex words will be more decomposable if they are less token-frequent than their bases and vice versa – the internal structure tends to become obscure as complex words become more frequent than their bases. For this reason, working with token-rare *-er* nouns all but guarantees that their base verbs will be more token-frequent and that the complex words would be processed/assembled with due consideration of morphological components.

Finally, we choose advanced EFL learners because we expect them to be similar to native speakers in having some productive ability with CMs, but not similar enough to be able to produce novel NCM types. Native speakers have round-the-clock opportunity to update their *-er* repertoire. They easily acquire the conventional locative *sleeper* ('rail car'), but also produce locative innovations like *refresher* ('Starship bathroom (Star Wars)'). As non-native learners lack the privilege of "complete" exposure, we expect that their knowledge of NCMs only comes at advanced levels of learning. Even then, given NCMs' low type frequency (see Corpus Analysis), we expect that this knowledge is item-based, not (yet) generalizable to novel instances. Systematic avoidance of *-er* in expressing such meanings (on a productive test) and systematic non-acceptance of those meanings in *-er* nouns (on a recognition test) would suggest the entrenchment of *-er* as an exclusively agent/instrument suffix even in the advanced learner.

Additional linguistic factors worth considering for their impact on learning *-er* include: the one-to-many, the many-to-one form-meaning relationships in the five semantic domains, and L1 influence.

The form-meaning mapping in *-er* is not 1:1. Deverbal *-er* could cue five meanings, which should make it a challenge to learn. However, since *-er* overwhelmingly links to agents and instruments, this is almost a non-issue. If learners have not encountered many NCMs and made the corresponding generalizations, or learned them by explicit instruction, to them *-er* remains consistent in symbolizing agents and instruments. The comparatively rare patients, locations and causes may not be robust enough to weaken learners' experience with and learning of the two CMs (or to enter EFL textbooks). The 1:5 ratio thus reduces to the more learnable 1:2.

Conversely, *-er* faces competition in coding the five meanings. Exposure to word-formation alternatives implies smaller extent of learners' experience with *-er's* meaning associations and having to learn the forms' respective distributions.[2] We cannot afford to discuss *-er*'s rivals (see Lieber, 2004). We only note in passing that not all competitors are equal. Although agents and instruments have some competition from compounding and less so from conversion (conversion to verbs being more productive), the other rival suffixes (*-ist, -or, -ian, -ant/-ent*) operate in technical/scientific contexts. Patients, locations and causes apparently have more competition from alternative affixes and from outside word formation, but their rival affixes tend to have different distributions (*-er* patients are inanimate, *-ee* patients human (Lieber, 2004)). Still, the less frequent association of *-er* with causes, patients and locations coupled with whatever competition there is from alternative expression forms conspire to reduce learners' experience of *-er* as a locative, patient and causative suffix.

Powerful usage-based pressures also come from L1. The entrenchment of L1 constructions and forms may facilitate, but also impede learning the constructions and forms of a new language (Ellis, 2008; Lowie, 2005; MacWhinney, 2008). We should, therefore, consider L1 impact on Croatian learners' acquisition of *-er*. However, L1 influence will be disregarded. L1 influence at the level of derivational constructions (suffixes) is virtually impossible to disentangle from intralingual influence: agent and instrument are simultaneously the most type-frequent meanings of English *-er* and the only meanings sanctioned by Croatian counterparts (*-aš*, *-ač*, etc., Babić, 2002; Silić & Pranjković, 2007). This also applies to specific *-er* nouns in Croatian, which are agent or instrument borrowings from German or English (*šloser* Germ. 'locksmith'; *menadžer* Engl. 'manager') or agent or instrument derivations on foreign bases (*špediter* 'freight forwarder' < Germ. *Spedition* 'freight forwarding'). The only patient noun attested in Anić (2004), a comprehensive dictionary of Croatian, is the borrowing *bestseller.* We may add *tester* 'product sample,' an instrument-adjacent patient found in the context of cosmetics. The two cannot have built a new pattern. The limited polysemy of Croatian agentive suffixes is thus very likely co-responsible for the learners' observed preference for agents and instruments discussed later.

There are, of course, many additional linguistic (e.g. contextual diversity, salience, prototypicality, generality, redundancy, translation equivalence), cognitive (e.g., language aptitude, learning processes, memory, attention), psychological (e.g. personality, motivation, self-regulation, learning style, anxiety), and sociocultural factors (e.g. context, cultural background) (Divjak & Caldwell-Harris, 2015; Dörnyei & Ryan, 2015; Ellis, 2012) that may affect EFL learners' success with *-er*. Some research suggests that even if they acquire native-like representations of complex words, second language learners may not actually use those representations in online recognition of complex

[2] For space constraints we cannot discuss our participants' susceptibility to this rivalry.

words, preferring to rely on nonstructural and semantic information instead (cf. Yoonsang et al., 2020). Though important, the challenge of integrating all these factors into a unified explanatory framework (Ellis, 2012) is beyond the scope of this paper and remains to be addressed in future.

2. Theoretical background and corpus analysis

2.1. *Theoretical background*

Usage-based models of language (UBM) as defined in Cognitive Linguistics (Barlow & Kemmer, 2000; Taylor, 2002) form the theoretical framework of our study. As models of language that aim at psychological reality, UBMs have strong affinities with what many psycholinguists and language acquisition specialists have claimed about language acquisition and use (Bybee, 2010; Ellis, 2002a; Lowie, 2000; Lowie & Verspoor, 2004, Tomasello, 2000, 2003; Tyler, 2010). They are empirical, bottom-up and put the language user center-stage. They prioritize authentic language data in building hypotheses about users' cognitive representations of language. They mainly rely on large authentic corpora, which represent reasonably well the natural discourse from which native users build their linguistic representations. While a corpus representative of linguistic input in a FL learning context is bound to remain elusive (see below), this does not defeat the principle that FL learning, like L1 development, involves, in some measure, finding patterns in the input and making generalizations using general cognitive abilities like categorization, analogy, comparison, abstraction (Bybee, 2010; Langacker, 2000).

According to UBMs, a user's language consists of a dynamic network of symbolic constructions of different levels of complexity, entrenchment and schematicity. Frequency of exposure to linguistic units is vital to the building of such networks via entrenchment and pattern extraction. Repetition of forms (token frequency) leads to entrenchment of their memory representations; repetition of types (type frequency) may lead to pattern recognition and its entrenchment as an abstract schema (Bybee 2008).[3] The stronger the

[3] Admittedly, the existence of many types does not in itself guarantee schema productivity (Baayen & Lieber, 1991; Bybee, 1995: 434). The standard argument is that patterns featuring many types are not likely to be productive if their instantiating complex words are of high token frequency. Frequent use of complex words strengthens their memory representations. They are accessed holistically, which is why they do not contribute to schema productivity (cf. Bybee & Moder, 1983). The importance of relative frequency was commented on earlier in text. We believe neither of these is a problem for the claim that *-er* is in general a productive suffix. We do register it in some very token-frequent nouns (*teacher* N = 138 526), but also in many with no more than 50 tokens (N = 593/1315 or 45%). Also, only 73 of 1315 *-er* nouns (or 5.56%) are more frequent than their bases, and our test items are not among the 73 (see Method).

entrenchment of specific forms, including regular ones, the more accessible they become. Through repetition they solidify into "prefabricated chunks" which users simply select in a communicative event. Strong entrenchment of a schema makes it easily available for producing novel types. Instances and schemas co-exist in users' cognitive representations of grammar and become available for interpreting and creating novel (uses of) linguistic units[4] (Cruse & Croft, 2004; Langacker, 2000; Taylor, 2002).

How does this translate to our concerns? FL learning may differ from L1 acquisition, but, as already noted, from the theoretical standpoint of UBMs, they both involve input-based, frequency-sensitive learning. Therefore, we make the following predictions about Croatian learners' acquisition of *-er*:

- *Schema entrenchment: type frequency and productivity*. The presumed linguistic input (see Corpus Analysis) to which advanced Croatian learners are exposed ideally includes *-er* nouns denoting agents, instruments, locations, causes and patients. The more nouns learners encounter from each category, the more likely they are to notice the patterns and abstract corresponding *-er* schemas: [[PROCESS/…]- [AGENT/*er*]]; [[PROCESS/…]-[INSTRUMENT/*er*]]; [[PROCESS/…]-[PATIENT /*er*]]; [[PROCESS/…]-[LOCATION/*er*]]; [[PROCESS/…]-[CAUSE/*er*]].[5] As schema strength (entrenchment and productivity) depends to an extent on the number of distinct types in the input, learners may have *-er* schemas of lower and higher productivity.[6] Therefore, learners should be capable of interpreting/building novel *-er* nouns of the relevant types if the corresponding schemas are well-entrenched in their cognitive systems.
- *Entrenchment of instances: token frequency and independent storage*. Highly token-frequent *-er* nouns may have been stored by learners as lexically specific constructions, e.g. [[READ/*read*]-[AGENT/*er*]] or [[DINE/ *dine*]-[LOCATION/*er*]]. In fact, any polysemous *-er* noun may be stored as a network of such constructions, each pairing the word's phono-logical form with one of its meanings, e.g. [[READ/*read*]-[AGENT/*er*]], [[READ/*read*]-[INSTRUMENT/*er*]], [[READ/*read*]-[PATIENT/*er*]].[7] These constructions may vary in degree of entrenchment due to different token frequencies, with *reader*$_{ag}$ being most entrenched and *reader*$_{pat}$ 'an-

[4] Novel instances also emerge by analogy to specific entrenched units. Langacker (1987) sees no difference in kind between schema- and analogy-based categorization since both involve similarity assessment.

[5] These schematic constructions do not entail full compositionality; their semantic components are only access points to more elaborate composite representations.

[6] Other factors affecting schema productivity, which we cannot discuss here, are its semantic and phonological openness, pragmatic usefulness, etc. (Taylor, 2002).

[7] We will refer to the lexically specific constructions pairing a different meaning with the same phonological form, like the reader constructions above as different meanings of the word.

thology' least/not entrenched. If the lexically specific construction is not entrenched (enough) but learners have the corresponding schema, they may build the target word/interpretation by accessing the schema. If no schema is available, they fail. All in all, lexically specific -*er* constructions may co-exist with their corresponding schemas in the learners' cognitive representation of -*er* if sufficiently token-frequent, but may also be autonomous if learners have not experienced a sufficient number of relevant types to make the generalization.

2.2. *Corpus analysis*

Our usage-based orientation requires establishing type and token frequencies of -*er*'s different meanings. A high type frequency of agents and instruments could be taken for granted. However, only a corpus analysis can supply token frequencies of specific -*er* nouns and the relative token frequencies of their co-existing meanings (Gries, 2008).

In the EFL context, such frequency data should ideally come from examining the *actual* learner input. However, accessing and quantifying the history of FL learners exposure may be an impossible feat (Ellis & Schmidt, 1997; Gries, 2008). We had to make concessions and decided to perform our corpus analysis on native language data found in the Corpus of Contemporary American English (COCA). We expected that the type and token frequencies thus obtained would give us some idea of what to expect of advanced learners for whom such naturalistic data represents a non-negligible portion of linguistic input.[8]

Database. Our database for the semantic and quantitative analysis is a list of deverbal -*er* suffixations from COCA.[9] They were not obtained automatically since the 7400 -*er* nouns that COCA automatically delivered had to be manually cleaned of simple nouns in -*er* (*lavender*, *mother*, etc.), -*er* suffixations that could not be interpreted as deverbal,[10] and -*er* suffixations with

[8] Film and television subtitles arguably better represent everyday language exposure than corpora based on written sources (Brysbaert & New, 2009). While our student-participants almost invariably claimed substantial exposure to the more spontaneous English (via TV, the Internet etc.), they also read many literary and academic texts for their coursework, which is why we went with the more balanced COCA.

[9] Our corpus analysis was performed when COCA numbered 450 million words.

[10] This was complicated by conversion. It is easy to assume a base noun in *hatter* 'one who produces hats', but not in *panicker*: 'a person who often feels panic' or 'a person who panics a lot'. We treated as deverbal all nouns whose apparent nominal base can be interpreted as verbal, cf. *panicker*. Since denominal nouns are semantically similar to deverbal (cf. Lieber, 2004: 18), we do not think their inclusion would change our corpus results.

compound bases.[11] The final database numbered 1315 nouns, after any misspelt, alternative orthographic/morphological forms were joined under their respective lemmas.

Semantic analysis. All nouns in our database were analyzed into five semantic categories. The categories were data-driven and reflect the *function* of the nouns' referents as agents, etc. in the underlying events (events coded by base verb), not their ontology (people, vehicles, etc.). There is psycholinguistic evidence that speakers are sensitive to entities' roles in larger relational structures rather than (only) literal identity when seeking similarities or analogies (Gentner & Markman, 1997; Gentner & Medina, 1998). For instance, we categorized vehicles as INSTRUMENTS when construable as aiding AGENTS in carrying out underlying events (*dredger* 'an instrument for dredging'), but as agent-like when such interpretation was excluded (e.g. *guzzler* 'a vehicle that guzzles too much gas' rather than '*an instrument for guzzling gas').

Our categories carry traditional semantic role labels. We do not subscribe to any theory of semantic roles since none has proved entirely satisfactory, especially in the face of massive data. True to the spirit of cognitive linguistics, though, we view them as flexible categories of experience, implying there is/are:

- no short, definitive list of roles/categories; their inventory can always be refined and updated based on finer-grained analysis
- no clean boundaries; the categories exhibit prototype structure and degrees of membership must be assumed.

AGENTS AND OTHER AGENT-LIKE ENTITIES: prototypical AGENTS are volitional, sentient humans responsible for initiating underlying events. Less prototypical are involuntary AGENTS like *yawner*, supernatural beings (*tempter* - devil, *forgiver* - God), animals (*hunters, breeders*) and plants (*rooters, seeders*). The category also includes energetic, nonvolitional, nonsentient inanimates like forces of nature/universe: e.g. *combers* (waves), *bringers* of life (rivers). They are *conceptualized* as effecting underlying events even if objective instigators exist earlier in the cause-effect chain: *combers* are caused by wind, quakes and landslides, but the wave is doing the 'combing'. Such inanimates are different from many INSTRUMENTS since they are *conceptualized* as carrying out the event by themselves, not as some AGENT'S 'tool'; e.g. *stabber* is INSTRUMENT when it means 'knife,' but AGENT when it means a type of pain.

[11] Many compound *-er* suffixations would be easily interpretable since their bases include verbs and are semantically similar to simple suffixations. We excluded them for consistency. COCA automatically delivered only compound-base suffixations with solid and hyphenated spelling. Including them would require also including the many open-spelling compounds (which are hard to distinguish from free phrases) and would make the corpus analysis inordinately complicated.

AGENTS, INANIMATE AGENT-LIKE ENTITIES and some INSTRUMENTS (those that need little human energy input, e.g. biological or chemical substances) are similar since they all 'do' what the verb implies. *Writers* write, *combers* metaphorically 'comb' sea surface, *repellers* repel insects etc. They differ in what is implied on top of their basic processual nature. AGENTS 'do' as willful, purpose-driven humans, AGENT-LIKE INANIMATES 'do' as inherently energetic forces; some INSTRUMENTS similarly 'do,' but only in serving specific human goals. [12]

INSTRUMENT: prototypically physical objects used by AGENTS to carry out underlying events, e.g. man-made artifacts (*stabber, crumber*), vehicles (*dredger, seeder*) etc. INSTRUMENTS are mostly incapable of sustaining the event without AGENT's more or less continuous energy input. However, this varies; cf. hand-held tools like *stabber* vs. computer programs like *uploader* where much is beyond AGENT's control after the initial mouse click. INSTRUMENTS 'interfere' with the semantic types described below, which makes all these categories non-homogenous.

The remaining three categories are treated as non-central due to their low type frequency. They feature prototypical and non-prototypical members conceptually close to instruments or agents (cf. Ryder, 1991). Such ambiguities were noted, and our corpus analysis was performed with conservative (only prototypical instances) and liberal (prototypical and non-prototypical instances) counts. Since under either analysis agents and instruments proved dominant, we only report conservative results (Appendix A).

LOCATION: prototypically LOCATIVE SETTINGS, i.e. places where underlying events unfold; e.g. *diner, sleeper*. However, there are many LOCATIVE OBJECTS, i.e. objects construable as the locus in/with respect to which the underlying event unfolds. This includes container-like setting objects with real or fictive boundaries, e.g. *fermenter* 'a container *in* which fermentation occurs' or *recliner* 'a piece of furniture *in* which one reclines.' Other objects function as the goal, path or source of action: e.g. *kneeler* 'an object *onto* which one kneels', *swiper* 'an object *through* which one swipes credit cards,' *gusher* 'a blood vessel *from* which blood gushes.' Most locative objects are interpretable as instruments (Luschützky & Rainer, 2013: 1302-1303): a *kneeler* is also 'an object *used* for kneeling,' a *swiper* is also 'an object *used* to swipe credit cards.' Source locatives may invite agentive interpretations, e.g. a *gusher* as 'a vessel which gushes blood.'

PATIENT: prototypically entities undergoing state change under AGENT's energetic action, e.g. *slicer* 'tomato for slicing.' PATIENTS also include entities

[12] Cf. Van Valin and Wilkins's (1996) thematic role *effector*, which underlies agents, natural forces and some instruments (dynamic participants doing some action, unmarked for volition and control).

construable as coming into being (Fillmore, 1968), e.g. *stacker* 'a burger with layers of meat etc. stacked up.' Some patients undergo change in apparently single-argument events, e.g. *bleeder* 'one who bleeds.' There are also instrument-adjacent PATIENTS. *Dipper* 'food, usually bread or vegetables one dips into a sauce' is arguably a true patient, but *scratcher* 'cats' scratching pole' is simultaneously 'an object that is scratched' and 'an object *used for* scratching (by cats).'

CAUSE: an entity or event causing the underlying event, e.g. *screamer* 'drugs which cause someone to scream.'[13] Prototypically, they involve complex events, consisting of the causing event (e.g. drugs affecting brain's chemistry) and the caused event (user screaming). One element from each event is invoked/explicitly coded: the causer from the causing event, i.e. drugs as the referent of *-er*, and the causee's reaction from the caused event, i.e. screaming as the base verb. The main participants of the causing and the caused event are distinct: drugs vs. user, and two events, though related, are quite distinct: drugs chemically interfering with brain vs. user screaming. Notice the ungrammaticality of '*drugs that scream someone.' The conceptual distinctness between the two events is gradient and may be blurred, which gives us less prototypical causes, e.g. *fader* 'UV rays fade hair or cause hair to fade'. The interpretation of UV rays as causes or more immediate agents depends on the perceived size of the conceptual gap between UV ray-specific chemical activity and hair-specific chemical reaction and which set of chemical processes strictly speaking qualifies as *fading*. However, it is highly unlikely that language users bother with such detail. They may respond to syntactic contexts in subconsciously categorizing particular examples, whereby they may perceive syntactically simpler causes like *fader* as agents. Also, many causes verge on instruments, e.g. *softener* 'chemicals which cause fabric to soften/which soften fabric/which *are used* to soften fabric.' Finally, the category CAUSE also allows *facilitative* interpretations (Talmy, 2000a,b): *socializers* 'parents who enable the socialization of their children.'

Quantitative analysis. We analyzed the overall token frequencies of nouns in our database and, for polysemous nouns, the token frequencies of their respective meanings (the latter given only for test items in Table 1). The overall token frequencies of the 1315 nouns ranged between 138526 and 1 tokens. We analyzed 500-token samples or all occurrences of nouns occurring less than 500 times. To be able to address our full set of hypotheses, we sorted the nouns into two frequency classes: high-frequency nouns (> 1000

---

[13] In support of the category CAUSE, Panther and Thornburg claim "a groaner does not 'groan you' in the same sense as a thriller 'thrills you.' Groaner denotes an event, often a bad joke, which *makes* the experiencer groan" (2002: 300).

tokens) and low-frequency nouns (<100 tokens).[14] As for type frequencies, our quantitative analysis confirmed the dominance of agents and instruments, lending support to our predictions stated earlier (Appendix A).

3. Hypotheses

Our general hypothesis was as follows: since *-er* nouns of semantic types that are type-rare in a native language corpus are even less likely to be available to EFL learners, advanced EFL learners have (a) item-based knowledge of high-frequency nouns in their dominant[15] NCMs, little or no knowledge of low-frequency nouns in NCMs and no corresponding schemas to interpret or build novel NCMs; (b) knowledge of high-frequency nouns in their dominant CMs *and* corresponding central schemas to interpret or build novel agents and instruments.

These predictions were empirically verified in a large-scale study involving advanced Croatian EFL learners. However, due to the volume of data, we have to be selective. In this paper we present results on two specific hypotheses concerning low-frequency nouns and only refer to some of the remaining (here unreported) hypotheses when necessary to contextualize aspects of the analyses presented.

The specific hypotheses concerning low-frequency nouns predict, among others, that participants would always perform better on CMs than NCMs, regardless of which meaning of the polysemous nouns is corpus-dominant (see Method), because participants have CM schemas, but no NCM schemas to fill expected gaps in their lexical knowledge of these nouns.

Hypothesis 1 (H1) concerns low-frequency, central meaning dominant nouns (LF+CDs): *spitter*: A>P, *fryer*: I>P, *sipper*: A>P, *boozer*: A>L, *refresher*: A>L, *nibbler*: A>P, *laugher*:–A>C, *weeper*: A>C. Hypothesis 2 (H2) concerns low-frequency, non-central meaning dominant nouns (LF+NCDs): *loaner*: P>A; *stuffer*: P>A; *crapper*: L>A; *shitter*: L>A; *groaner*: C>A; *yawner*: C>A; *snoozer*: C>A.

We analyzed the results from two perspectives. Since our target nouns are polysemous between CMs and NCMs, we first compared *relative* success with the co-existing meanings to establish whether participants were more correct on CMs or NCMs (H1(a) and H2(a)). Since being 'better on' can mean 'more good at', but also 'less bad at', we also considered participants' answers on each meaning *independently*, by measuring their correctness scores

[14] There is no natural cut-off point between items that should be treated as frequent and those that are less (or not) so. Any decision on where to draw this line is arbitrary (Gries 2008).

[15] "Dominant" means more frequent relative to the other meanings of the same noun, but should not be understood as a mathematically precise concept.

up against an external standard (participants' scores on item-familiar HF words like *reader*$_{ag}$ or *drawer*$_{pat}$, analyzed under some of the here unreported hypotheses) (H1(b, c) and H2 (b, c)).

- **H1(a)** *In LF+CDs, participants are **more correct** with dominant CMs than subordinate NCMs*

  We hypothesized that participants would be consistently better with the dominant CMs of LF+CDs than their subordinate NCMs. Given the nouns' LF, we do not expect that quantitative advantage of CMs over NCMs matters or guarantees the former's memory representation (lexical knowledge). But we do expect that this advantage is largely due to CM schemas (and non-existence of NCMs schemas), which step in where lexical knowledge is absent.

- **H1(b)** *Participants are **as correct** with CMs of LF+CDs as they are with CMs of HF+CDs*

  We assumed that participants would be equally good with CMs of LF+CDs and CMs of well-known HF+CDs since they can tap two resources to catch up with HF nouns: lexically entrenched instances (if any) and CM schemas.

- **H1(c)** *Participants are **less correct** with NCMs of LF+CDs than with NCMs of HF+NCDs*

  We expected that participants were sufficiently exposed to the NCMs of HF+NCDs to have committed them to memory. If they had any knowledge of the NCMs of LF+CDs, it would be exceptional at best. There being no NCM schemas, NCMs can only be known holistically.

- **H2(a)**: *In LF+NCDs, participants are **less correct** with dominant NCMs than subordinate CMs*

  This hypothesis makes almost identical claims as H1 about mirror-image data—these LF nouns are NCD and their CMs are subordinate. This quantitative asymmetry, as already said, may matter little given the nouns' overall LF, but here we allowed that participants may know at least some nouns in their dominant NCMs. If participants proved better on CMs nonetheless, the case for CM schemas would be even stronger.

- **H2(b)** *Participants are **less correct** with NCMs of LF+NCDs than NCMs of HF+NCDs*

  This follows from the assumption of entrenched lexical knowledge of most/all HF nouns and at best accidental lexical knowledge of some LF nouns in their dominant NCMs.

- **H2(c)** *Participants are **as correct** with CMs of LF+NCDs as they are with CMs of HF+CDs*

This follows from the expectation that central schemas would make up for lack of lexical experience with CMs of LF+CDs.

4. Method

4.1. *Selection of test items and fillers*

We decided to work with real but, in this study, rare words, rather than nonce-words since we did not want nonsense bases drawing undue attention to themselves (Nagy et al., 1993). Because we used existing words, lexical polysemy was inevitable; there is virtually no NCM noun that cannot also be interpreted as an agent or instrument. There is some risk of perplexing participants by prompting them for unfamiliar meanings of familiar words, however, we think the risk is not substantial. Our participants are EFL learners for whom the corpus-rare nouns, despite being base-familiar, may not be part of the conventional, well-entrenched vocabulary. Thus, their coexisting meanings should not stand in each other's way.

Twenty seven out of 1315 nouns were selected for testing our full set of hypotheses. We used the following criteria:

1. **prototypicality**: nouns should be typical representatives of the five categories, especially the NCMs
2. **polysemy**: nouns should feature a CM and a NCM
3. **overall token frequency**: there should be nouns with high frequency (HF) and low frequency (LF)
4. **relative frequency of meanings**: relative to overall token frequency, nouns should be either central meaning dominant (CD) or non-central meaning dominant (NCD)
5. **transparency of verb base**

Since NCMs are infrequent, it was difficult to honor the criteria of prototypicality and verb base transparency, but we adhered strictly to the remaining three. Of these, criterion 3 is irrelevant for the 15 LF nouns whose analysis is reported here. The 15 nouns feature various combinations of central (A and I) and non-central meanings (P, L, C). Table 1 shows their corpus-attested meanings and token frequencies.

A few atypical meanings call for a brief description: *spitter*$_{pat}$ = sour apples, (baseball) ball with spit on it; *fryer*$_{pat}$ = food (chicken) suitable for frying; *sipper*$_{pat}$ = drinks to be consumed slowly; *boozer*$_{loc}$ = a pub, bar; *refresher*$_{loc}$, *crapper*$_{loc}$, *shitter*$_{loc}$ = a restroom; *nibbler*$_{pat}$ = snacks; *laugher*$_{caus}$, *weeper*$_{caus}$, *snoozer*$_{caus}$, *yawner*$_{caus}$, *groaner*$_{caus}$ = something causing laughter, sadness, boredom, displeasure; *loaner*$_{pat}$ = something loaned as replacement; *stuffer*$_{pat}$ =

brochures enclosed with regular mail, or something, like tomatoes, suitable for stuffing.

Table 1: Test items

Central meaning dominant			Noncentral meaning dominant		
Noun	**Overall token freq**	**Relative[16] Token freq**	**Noun**	**Overall token freq**	**Relative Token freq**
spitter** (A, P)	48	A = 31* P = 16* I = 1	loaner (P, A)	54	P = 49* A = 5*
fryer (I, P)	60	I = 45* P = 12* A = 3	stuffer** (P, A)	20	P = 18* A = 2*
sipper (A, P)	46	A = 28* P = 10* I = 8	crapper (L, A)	49	L = 43* A = 6*
boozer (A, L)	55	A = 53* L = 2*	shitter (L, A)	30	L = 28* I = 1 A = 1*
refresher (A, L)	101	A = 81* I = 19 L = 1*	groaner (C, A)	8	C = 7* A = 1*
nibbler (A, P)	30	A = 17* I = 10 P = 3*	yawner (C, A)	17	C = 17*
laugher** (A, C)	55	A = 33* C = 22*	snoozer (C, A)	27	C = 17* A = 10*
weeper (A, C)	48	A = 25* C = 23*			

A = agent, I = instrument, P = patient, L = location, C = cause
* meanings selected for analysis
** base verbs on COCA's 5000-word frequency list

A few more comments are necessary. *Refresher* had one occurrence over the LF limit, but was included because it was hard to find eligible LF nouns. There is a noun with only one meaning attested in COCA, *yawner*. Since it can also mean 'a person who yawns' (as easily verified in any dictionary), we supplied the corpus-unattested agentive meaning. Two non-prototypical instances are *crapper*$_{loc}$ and *shitter*$_{loc}$, which meant prototypical location 'restroom' in 16 out of the 43 'locative' cases and 8 out of 28 'locative' cases respectively, the remaining locatives were metonymically associated instru-

[16] Here relative token frequency means the token frequencies of different meanings of the same -er noun (not how it is interpreted in Hay, 2001). But it may be useful to reiterate here that all the test nouns were less token-frequent than their base verbs.

ment-locations 'toilet bowl', or in *shitter* holes dug in the ground with the same function. Since the prototypical locatives still outweighed the agentive 'person defecating' or the figurative version of it, these nouns were included.

Verb bases are largely transparent. Although only 3 (marked ** in Table 1) are found on COCA's 5000-word frequency list,[17] the remaining 12 concern basic human physiology and mundane activities, which makes them eligible as basic vocabulary items.

We describe the rest of methodology as it pertains to the full set of analyzed data. We chose 3.27 fillers per target for the recognition task (T1R) and 2.4 fillers per target for the productive task (T2P). Given the total number of definitions/sentences, we minimized as far as possible the number of fillers without risking revealing our research goals. T1R has fewer fillers than T2P since we assumed it would be harder to detect our intention where respondents had to produce *-er* targets from scratch. The fillers were similar to targets. Formally, they were noun suffixations in *-ation*, *-ance*/*-ence*, *-ity*, *-ing*, *-ment*, *-ling*, *-ee*, *-dom*, etc. Semantically, they were also polysemous. Since *-er* words have mainly concrete meanings, to blur the distinction between targets and fillers we incorporated fillers with concrete suffixes, e.g. *-ette*, *-ling*, *-age*, *-ery* and in T1R added some invented locative, patient, cause, agent, instrument meanings (e.g. fn. 18).

4.2. *Participants*

The a-priori power analysis indicated that to obtain a medium effect size of .5 and a desired power (1-β error probability) of .8 using an α level of .05, the total sample size required is 34. The critical *t* is 2.03. With the same assumptions, to obtain a large effect size of .8, the total sample size required is 15. The critical *t* is 2.1.

The participants were 131 first-, second-, third- and fourth-year English majors at a Croatian public university. They form quite a homogenous group in terms of their characteristics as language learners (see Appendix B; the "Demographics Questionnaire" is available on request). They have been learning English for 12.61 years (*SD* = 1.92), although they started learning at different ages. Most come from higher-quality secondary schools (85%), but only five (3.9%) visited an English-speaking region. They all had to qualify for enrolment in the English program, and claim to be highly motivated. Since authentic English, especially informal American, is becoming ever more present in Croatian public and private discourse (TV, the Internet, social networks, advertising, etc.), it is unsurprising that participants attribute much of their current English competence to implicit non-institutional

[17] http://www.wordfrequency.info/free.asp?s=y.

learning (the particular sources of exposure may differ, e.g., reading, TV, surfing, gaming, etc.). However, as English majors, they also read literature and academic texts for their course-work. Participants were randomly assigned to Test 1 (n = 71) and Test 2 (n = 60) and completed the test and the demographic questionnaire under a numerical, computer-assigned code.

4.3. *Test materials and procedure*

Both tests were prepared and administered using specially designed in-house software. The first test (T1R 'Test 1: Receptive') was a yes-no task where participants saw a series of words with their suggested meanings ("Test 1: yes-no task"; available on request). Their task was to indicate, by pressing a key on the keyboard, whether they thought the suggested meaning was correct/plausible or not. There was no key for 'I don't know'. The participants were instructed to try to guess if they did not have an immediate answer. The test included 235 meanings, of which 55 were the meanings of 27 *-er* nouns. The remaining 180 were fillers, with 34 pseudo-meanings.[18] Fillers and *-er* nouns were presented in a pseudo-randomized order to prevent formal or semantic association between consecutive items.

Meaning descriptions were carefully constructed based on those provided in the *Oxford Advanced Learner's Dictionary* (where available) so that they are equal in length (20 ± 1 syllables), free of complex constructions and unusual vocabulary and feature the base verbs to maximize transparency. The stimuli (word followed by one of its meanings) were presented one by one. Each stimulus remained displayed until the participant pressed the key. Participants could not return to change their answers. Cf.:

> ***sipper***
> 'a drink that is meant to be consumed slowly, by taking small mouthfuls, sip by sip'

Two keyboard keys were assigned the 'yes' and 'no' functions and marked with color stickers - green for 'yes' (letter key "P" at the right end of keyboard), red for 'no' (letter key "W" at the left end of keyboard). The space bar was used to call the next item onto screen.

Before the test, participants were told that we were interested in English words with more meanings and wanted to learn which meanings they knew. They were instructed on how to complete the task (oral and PowerPoint presentation), the functions of the keys were explained and were given a test round of 10 simple nouns. They completed the test on the same day during regular class time. The test took on average 30 minutes.

[18] E.g. for sharkling we added the nonsensical 'a cage used by marine biologist divers to protect them from shark attacks.'

After the test, participants completed a demographic questionnaire probing their age, language learning experience, exposure to English and motivation.

Test 2 (T2P 'Test 2: Productive') was a gap-fill test that included 187 sentences with gaps, 55 of which targeted the 27 *-er* nouns, the rest were fillers (132) ("Task 2: gap-fill test"; available on request). A verb was given in brackets at the end of the sentence. Participants had to build a noun (1 word only) from this verb and use it to fill the gap. The target and filler items were presented in a pseudo-randomized order.

Test sentences were carefully constructed so that they represent as closely as possible typical contexts for the target nouns. They were equal in length (20 ± 1 syllables), free of complex constructions and unusual vocabulary, and mostly featured the gaps near the end of the sentences. The sentences were presented one by one and participants could not return to change their answers. Cf.

His car was damaged in an accident; the car he's driving now is a ______________. (loan)

Before the test, participants were told that we were testing their knowledge of English words. They received oral and PowerPoint instructions. They completed the test on the same day during regular class time. On average the test took 45 minutes. After the test, participants completed the demographic questionnaire.

4.4. *Scoring*

On T1R we scored as correct or incorrect participants' acceptance or non-acceptance of the proposed meanings. On T2P, responses were scored as correct if the gaps were filled with target *-er* forms. Responses were coded as incorrect if semantically/grammatically unacceptable non-target words were supplied. With some nouns participants provided semantically potentially acceptable non-target responses, which were not coded as incorrect, but were treated as missing values. There were also some 'real' missing values, both circumstances leading to *n*s < 60 (total number of T2P participants).

4.5. *Analyses*

McNemar test with binomial distribution was used to assess how CMs and NCMs compare in terms of their relative proportions of correct vs. incorrect answers. Due to unequal distribution of *-er* nouns across groups, group variables were calculated as means of correct answer proportions. After outliers were deleted skew and kurtosis indices were within acceptable limits (SI < 3, KI < 10; cf. Kline 2010, p. 63) allowing for parametric tests for comparisons of related means to be conducted. To determine effect size Cohen's *d* was calculated applying a formula correcting for dependence between means, using Morris and DeShon's (2002) equation 8. Magnitudes of *d* were interpreted using criteria initially suggested by Cohen (1988) and expanded by Sawilowsky (2009).

5. Results and discussion

5.1. *Results (H1)*

All three subhypotheses of H1 were confirmed on T1R, and one was rejected on T2P. T1R group results confirmed H1(a). Paired samples t-test[19] showed that participants were significantly more correct at responding to dominant CMs of LF+CDs (M = .76, SD = .19) than their subordinate NCMs (M = .45, SD = .22), 95% CI [.25, .36], $t(70)$ = 11.71, p < .001, d = 1.4. although they did positively respond to some NCMs.

Table 2: Relative success with CMs vs. NCMs at individual level (H1, T1R)

Noun	Correctness (%)		McNemar (sig.)
	CM	NCM	
sipper	51 (71.8)	55 (77.5)	.557
nibbler	60 (84.5)	34 (47.9)	.000***
boozer	59 (83.1)	23 (32.4)	.000***
refresher	38 (53.5)	27 (38)	.080
weeper	59 (83.1)	52 (73.2)	.167
laugher	37 (52.1)	43 (60.6)	.286
fryer	69 (97.2)	13 (18.3)	.000***
spitter	60 (84.5)	12 (16.9)	.000***

*p < .05, **p < .01, ***p < .001; n = 71

McNemar test showed that the proportion of correct answers for LF+CDs in their CMs differed significantly from that for the same words in their NCMs in four cases (Table 2). *Nibbler*, *boozer*, *fryer* and *spitter* showed a significantly larger proportion of correct than incorrect answers on CMs than NCMs. Participants were also more correct than incorrect in assessing the plausibility of CM definitions of each of the four words.

[19] Alpha level of .05 was used for all statistical tests.

The remaining four nouns: *sipper, weeper, laugher* and *refresher* show no significant difference between CMs and NCMs. Considering this together with the proportion of correct vs. incorrect responses on each of their meanings independently, we see that participants were *equally good* on both meanings of *sipper* and *weeper* and *equally not-that-good* on both meanings of *laugher* and *refresher*.

We next compared participants' performance on LF+CDs with their performance on HF nouns used as standards of comparison.

According to H1(b), participants should be equally good with CMs of LF+CDs as with CMs of presumably well-known HF+CDs. The HF+CDs, analyzed under one of here unreported hypotheses, include *mixer*$_{inst}$, *reader*$_{ag}$, *sleeper*$_{ag}$, *smoker*$_{ag}$, *sucker*$_{ag}$, *keeper*$_{ag}$, *thinker*$_{ag}$. Group comparison of T1R results revealed that participants were equally correct on HF+CDs (M = .80, SD = .14) and LF+CDs (M = .76, SD = .19), 95% CI [-.078, .001], $t(70)$ = -1.95, P = .056, d = -.24.

According to H1(c), participants should be less correct with NCMs of LF+CDs than NCMs of the presumably well-known HF+NCDs. The HF+NCDs, also analyzed under one of here unreported hypotheses, include *diner*$_{loc}$, *drawer*$_{pat}$, *layer*$_{pat}$, *sticker*$_{pat}$, *trailer*$_{pat}$. Indeed, participants were significantly more correct at responding to dominant NCMs of HF+NCDs (M = .89, SD = .14) than NCMs of LF+CDs (M = .41, SD = .22), 95% CI [-.55, -.41], $t(55)$ = -13.38, p < .001, d = -1.827.

T2P group results also confirmed H1(a). Paired samples t-test showed that participants were significantly more successful in supplying the correct LF+CDs when the target was their CM (M = .72, SD = .15) than their subordinate NCM (M = .31, SD = .24), 95% CI [.36, .47], $t(57)$ = 14.95, p < .001, d = 2.1. As Table 3 shows, in all cases but one the proportion of correct to incorrect answers was in favour of CMs.

Table 3: Relative success with CMs vs. NCMs at individual level (H1, T2P)

Target noun	*n*	Correctness (%)		McNemar (sig.)
		CM	NCM	
sipper	60	52 (86.7)	3 (5)	.000***
nibbler	24	22 (91.7)	8 (33.3)	.000***
boozer	59	49 (83.1)	16 (27.1)	.000***
refresher	60	4 (6.7)	25 (41.7)	.000***
weeper	60	59 (98.3)	40 (66.7)	.000***
laugher	37	16 (43.2)	3 (18.8)	.000***
fryer	54	38 (70.4)	23 (42.6)	.001***
spitter	60	53 (88.3)	7 (11.7)	.000***

**p* < .05, ***p* < .01, ****p* < .001

Next, we compared T2P scores on CMs of LF+CDs and HF+CDs (H1(b)), then T2P scores on NCMs of LF+CDs and HF+NCDs (H1(c)).

According to H1(b), participants should be as correct with CMs of LF+CDs as with CMs of HF+CDs. However, we found a statistically significant difference between CMs of LF+CDs (M = .75, SD = .13) and HF+CDs (M = .96, SD = .06), 95% CI [-.25, -.17], $t(46) = -10.94$, $p < .001$, $d = -1.76$, in favour of the latter.

H1(c) was confirmed. Paired samples t-test revealed a statistically significant difference between NCMs of LF+NCDs (M = .32, SD = .24) and HF+NCDs (M = .83, SD = .20), 95% CI [-.56, -.44], $t(54)=-16.79$, $p < .001$, d = -2.29

Discussion of results (H1). Both tests confirmed H1(a), i.e. our expectation of better performance on dominant CMs than subordinate NCMs. Still, some LF agents and instruments were consistently misidentified as unacceptable or missed. Participants also did well on some NCMs although, given the low group score on NCMs, this knowledge is probably isolated. These findings are more interesting when compared between T1R and T2P.

CMs had a slim advantage (4 non-significant results) over NCMs on the more spontaneous T1R. This advantage increased when participants had to produce the target meanings in T2P. This is not because of dramatic improvement on CMs. The CM mean score was even slightly lower on T2P than T1R, suggesting that some CMs were systematically missed and non-target/incorrect forms were provided at least as often as the targets. The change is due to dramatically worse performance on NCMs when participants were left to their own devices. Three non-significant results from T1R turned in favour of CMs: with *sipper*$_{pat}$ and *laugher*$_{caus}$ dramatically so; the proportion of correct to incorrect answers on the NCMs plummeted. In *weeper*$_{caus,}$ the proportion of correct answers dropped and turned the tables in favour of *weeper*$_{ag.}$ Still, this it is the only NCM that kept the healthy proportion of correct to incorrect answers from T1R. The only meaning where the change went in the opposite direction, in favour of NCMs, was *refresher.* This is not because participants were particularly good with *refresher*$_{loc}$, but because they were very bad with *refresher*$_{ag}$. Let us examine closely the underperformers among CMs because they defy our correctness expectation and the overperformers among NCMs since they may not be chance.

CM underperformers. Two nouns proved particularly thorny. One is *refresher*$_{ag}$ as ‘lesson that refreshes one’s memory’. Granted, *refresher*$_{ag}$ is not a prototypical agent, but the only alternative type the noun arguably instantiates is instrument ‘lesson used to?/designed to refresh one’s memory’, another CM category. We attribute the poor results to interference from *refreshment*, a noun presumably more entrenched and similar enough to the target to have been misidentified as the 'host' of the proposed meaning/target. We cannot know for sure that this caused failure in T1R. But we can be fairly certain since participants also failed in T2P, where 48 of 60 participants supplied this infelicitous non-target.

Another poor performer was *laugher*$_{ag}$. Though statistically not worse than *laugher*$_{caus}$ on T1R, *laugher*$_{ag}$ was just as often misidentified as unacceptable as it was found acceptable. Despite its low frequency, weak performance on *laugher*$_{ag}$ would be surprising if we did not learn after the test sessions that participants had confused the form for a misspelt *laughter,* thinking it was "a trick question". Regardless, strong entrenchment of the noun *laughter* must have led to phonological misidentification of *laugher*$_{caus}$ as *laughter* and to rejecting the meaning 'a person laughing/who laughs a lot', so obviously incompatible with the assumed form (abstract vs. human). It is not certain that the same phonetic misidentification contributed to poor success on *laugher*$_{caus.}$ If there had been any phonetic confusion at all, it would arguably have *promoted* the identification of *laugher*$_{caus}$ as correct since *laughter* and *laugher* as 'cause of laugh<u>t</u>er' are semantically closer (both fairly abstract). Of course, whatever success participants did have with *laugher*$_{caus}$ may have nothing to do with any phonological accident, but with the fact that they simply knew the noun *laugher*$_{caus}$. T2P participants also performed rather poorly on *laugher*$_{ag}$, but since they failed much more on the NCM than the CM (cf. T1R scores where participants were better on *laugher*$_{caus}$), the CM took upper-hand in T2P. So *laugher*$_{ag}$, like *refresher*$_{loc}$, is an opportunistic winner. Simultaneously, *laugher*$_{ag}$, *refresher*$_{ag}$ and possibly *fryer*$_{inst}$ may have caused a lower mean value for CMs on T2P than T1R. The latter's drop in success on T2P may be due to entrenchment of *frying pan* in our EFL learners' repertoire, strong enough to have caused them to prefer the semantically imperfect compound.

NCM overperformers. These are not just nouns where McNemar showed no significant difference between the proportions of correct answers on the two meanings, but those where this is accompanied by a high level of correctness on the NCM. This is why on T1R *refresher*$_{loc}$ is not considered an 'overperformer' and *sipper*$_{pat}$, and *weeper*$_{caus}$ are. Of the two, only *weeper*$_{caus}$ showed consistency in participants also performing well on T2P. So, this is the only NCM noun among LF+CD nouns which is known consistently enough to be considered part of the network of well-entrenched, memorized *-er* nouns. We are reluctant to think the same of *sipper*$_{pat}$ purely on the evidence of its good performance on a test where there is a risk of guesswork. It seems that the low corpus frequency of *weeper* matters little. Participants had consistently good results on the noun, possibly suggesting lexical entrenchment, especially of its NCM. Whether or not a different, more youth-oriented corpus would show a higher token frequency of *weeper*$_{caus,}$ these results still suggest that factors other than objective frequencies also have a hand in learning. Different linguistic units/interpretations may have different degrees of subjective salience for different (groups of) speakers (Bley-Vroman, 2002; Bybee, 2007; Divjak & Caldwell-Harris, 2015).

All in all, although H1(a) was confirmed on both tasks, results also indicate that T2P may be more informative about what our participants actually know. With the exception of *weeper*$_{caus}$, T1R may have overestimated participants' knowledge of NCMs. On the more 'diagnostic' productive test, participants were only better on the NCM than CM in the one noun where they were poor on both, *refresher*$_{loc}$. On the other hand, with that one exception, participants were consistently significantly better on CMs than NCMs in both tasks, which in all cases but *laugher*$_{ag}$ implied they also knew the CMs well. As for the diagnostic value of T1R, if participants have a 50:50 risk of choosing wrong on T1R, the least bit of 'semantic noise' in the meanings to be assessed may be enough to throw them off the right decision (this could perhaps explain why no agents on T1R scored 100%). If, however, participants perform better when building target nouns from the ground up, then a significantly higher success rate with unfamiliar CMs compared to NCMs must be a sign that this device is a strong central schema, if not well-entrenched item-based knowledge.

T1R confirmed and T2P disproved H1(b), i.e. our assumption that participants would perform equally well on CMs of all CD nouns regardless of frequency. Participants were equally correct at recognizing CMs of LF+CD nouns (T1R), but less correct at producing them (T2P). We suspect T1R may have confirmed H1(b) for the wrong reasons. Since H1(a) turned up some CM underperformers in T1R, equal results between LF and HF CMs may be due to imperfect performance on one HF agent. Participants scored poorly on HF *sucker*$_{ag}$ (25.4% correct vs. 74.6% incorrect), probably because we went with the literal definition 'a young baby or a young animal that still sucks the mother's breast or udder' over the probably more entrenched figurative 'a gullible or annoying person' or 'a person strongly attracted to a thing/person of a particular type'. The proposed meaning was arguably also blocked by *suckling,* a more entrenched animal agent noun. In T2P, results on CMs of LF+CDs could hardly match the perfect 100% scores on HF agents like *reader*$_{ag}$ and *smoker*$_{ag}$, which resulted in HF nouns taking over the lead. On T2P participants were significantly more correct than incorrect even with *sucker*$_{ag}$, but the sentence was more felicitous, featuring mosquitoes and vampires rather than unweaned humans/animals.

Both tests revealed, as expected, a statistically significant difference in correctness on NCMs between the two frequency groups in favour of HF+NCDs (H1(c)). There were no outstanding results, so we conclude that better performance on NCMs of HF words is due to strong entrenchment of most/all those words, as opposed to isolated knowledge of a LF noun or two, best candidate being *weeper*$_{caus}$.

5.2. *Results (H2)*

One subhypothesis of H2 was confirmed on T1R, and two were confirmed on T2P. H2(a), predicting weaker performance on dominant NCMs than subordinate CMs in LF+NCDs was rejected on T1R. There was no statistically significant difference between NCMs of LF+NCD (M = .59, SD = .23) and CMs of LF+NCD (M = .61, SD = .28), 95% CI [-.11, .06], $t(70)$ = -.53, p = .599, d = -.06.

McNemar test revealed three out of seven nouns where participants were more correct than incorrect on subordinate CMs than their NCMs (*groaner, loaner, stuffer*). Two nouns pushed the group score in the opposite direction (*shitter* and *crapper*) and two non-significant results (*yawner, snoozer*) secured it in the no-significance zone (Table 4).

Table 4: Relative success with NCMs vs. CMs at individual level (H2, T1R)

Noun	Correctness (%)		McNemar (sig.)
	NCM	CM	
groaner	41 (57.7)	57 (80.3)	.011***
yawner	51 (71.8)	51 (71.8)	1.00
snoozer	43 (60.6)	48 (67.6)	.487
shitter	58 (81.7)	30 (42.3)	.000***
crapper	50 (70.4)	29 (40.8)	.000***
loaner	29 (40.8)	51 (71.8)	.000***
stuffer	23 (32.4)	40 (56.3)	.002***

*$p \le .05$, **$p \le .01$, ***$p \le .001$

H2(b), predicting that participants should be less correct with NCMs of LF+NCDs than with NCMs of the same HF+NCDs listed in H1, was confirmed on T1R. The difference between NCMs of LF+NCDs (M = .57, SD = .23) and NCMs of HF+NCDs was statistically significant (M = .89, SD = .14), 95% CI [-.38, -.26], $t(55)$ = -10.65, p < .001, d = -1.5.

According to H2(c), participants should be as correct with CMs of LF+NCDs as they are with CMs of HF+CDs. Paired-samples t-test disproved H2(c) on T1R; there was a significant difference between CMs of HF+CDs (M = .80, SD = .14) and CMs of LF+NCDs (M = .61, SD = .28), 95% CI [-.25, -.12], $t(70)$ = -6.07, p < .001, d = -.8, in favour of the former.

As for T2P, group results confirm H2(a): participants were statistically significantly better at subordinate CMs of LF+NCDs (M = .8, SD = .19) than their dominant NCMs (M = .43, SD = .28), 95% CI [-.43, .31], $t(55)$ = -12.27, p < .001, d = -1.77.

Table 5: Relative success with NCMs vs. CMs at individual level (H2, T2P)

Target noun	n	Correctness (%)		McNemar (sig.)
		NCM	CM	
groaner	60	14 (23.3)	36 (60)	.000***
yawner	42	32 (76.2)	39 (92.9)	.039*
snoozer	29	17 (58.6)	25 (86.2)	.008**
shitter	58	31 (53.4)	51 (87.9)	.000***
crapper	58	39 (67.2)	38 (65.5)	1.00
loaner	57	15 (26.3)	42 (73.7)	.000***
stuffer	59	3 (5.1)	42 (71.2)	.000***

*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$

Individual comparisons revealed that with one non-significant result (*crapper*), participants were always more correct than incorrect on subordinate CMs than the dominant NCMs. This does not mean that they were bad with all NCMs or good with all CMs. Even in the non-significant *crapper,* participants were more correct than incorrect on each of the meanings observed independently (Table 5).

H2(b) was also confirmed on T2P: participants were statistically significantly less correct with NCMs of LF+NCDs (M = .44, SD = .27) than with NCMs of the same HF+NCDs listed in H1 (M = .83, SD = .2), 95% CI [-.44, -.32], $t(54)$ = -12.07, p < .001, d = -1.69.

H2(c) was rejected on T2P: participants were not as correct with CMs of LF+NCDs as with CMs of HF+CDs. Group comparison revealed a statistically significant difference between CMs of LF+NCDs (M = .83, SD = .18) and CMs of HF+CDs (M = .96, SD = .06), 95% CI [-.19, -.08], $t(46)$ = -5.2, p < .001, d = -0.86.

Discussion of results (H2). According to H2(a), we expected that the dominance of NCMs over CMs in LF+NCDs would not matter. Given the nouns' overall LF and likely existence of CM schemas, participants were expected to outperform on CMs. T2P confirmed this, but the CM advantage on T1R was not significant. Let us consider individual results.

T1R: In defence of participants' performance on CMs, they were significantly more correct than incorrect with four agentive nouns and in three cases there was no significant difference in the proportion of correct vs. incorrect answers (cf. Table 4). But this was insufficient to secure CMs' group advantage over NCMs. In several nouns, participants were equally good with NCMs (*yawner*$_{caus}$ and *snoozer*$_{caus}$) or better (*crapper*$_{loc}$, *shitter*$_{loc}$). We suspect this is not accidental. The former are semantically similar to *weeper*$_{caus}$ from H1. All five may be close to our student-participants: they are informal, strongly evaluative and expressive, which may boost their salience and compensate for corpus-rarity. With two vanilla-type nouns *stuffer*$_{pat}$ and *loaner*$_{pat}$ participants were far below this performance level. If any nouns in our database support Glynn's (2014: 15) observation that not all input is equal, it is these five.

T2P: The explanation for the group score was already hinted at above. Non-central *-er* targets were easily missed: e.g. participants did not use *loaner* for 'loaned car', but used participial paraphrase *loaned (one)*. On the other hand, they produced these same LF+NCDs in their CMs quite consistently, but did not score a perfect 7/7 either.

NCMs (T2P): Looking at each meaning independently in Table 5, we see that participants performed well on two of the four well-performing NCMs from T1R, viz. *yawner*$_{caus}$ and *crapper*$_{loc}$. In *shitter*$_{loc}$ and *snoozer*$_{caus}$ they were more correct than incorrect but below significance level. As for *snoozer*$_{caus}$, the most frequently provided non-target forms were conventional alternatives *snooze, snoozefest.* The acceptable alternative *yawn*$_{caus}$ was occasionally found among non-targets for *yawner*$_{caus}$. We presume the existence of well-entrenched alternatives for the causative meaning may have weakened participants' otherwise good results on *snoozer*$_{caus}$. We also believe, but leave for future research, that *snoozer*$_{caus}$ and *yawner*$_{caus}$ have some degree of entrenchment and, together with H1 noun *weeper*$_{caus}$, are forming a low-level analogy-based cluster in our participants' cognitive systems. As for *crapper*$_{loc}$ the prevalence of correct responses over incorrect ones was significant, but was non-significant with *shitter*$_{loc}$. The 'demise' of NCMs on T2P is most likely due to dramatically poorer results on *groaner*$_{caus}$, *loaner*$_{pat}$, *stuffer*$_{pat}$. Participants were both poor on those meanings and poorer on those meanings than the CMs of the same nouns (Table 5). If this indicates that they do not really know *groaner*$_{caus}$, *loaner*$_{pat}$, *stuffer*$_{pat}$ and lack the supporting schemas, their relatively better performance in T1R can be attributed to the more passive nature of the task (see General Discussion).

CMs (T2P): Better performance on CMs than NCMs is also partly due to significant improvement in CMs relative to T1R. The score is still not a perfect 7/7 since performance was suboptimal on *groaner*$_{ag}$ (worse than on T1R) and not so brilliant on *crapper*$_{ag}$ (though better than on T1R). We suspect that some participants struggled with the lexical meaning of *groan.* Their relative lack of success with *crapper*$_{ag}$ appears surprising at first, especially since the noun is practically synonymous with *shitter*$_{ag}$, in both its literal and figurative sense. Arguably, this might be due to stronger entrenchment of *shitter*$_{ag}$ as agent, cf. the conventional phrase "Don't shit a shitter". Also, participants were drawn to the more familiar, semantically compatible, but syntactically non-felicitous noun *crap* (11/20 incorrect responses). Namely, the sentence featured an indefinite article, expecting the countable figurative target *crapper*$_{ag}$.

Since we had different participants for the two tasks, we will only note in passing that, when each meaning was considered independently, participants were significantly better with CMs on T2P than T1R and significantly worse with NCMs on T2P than T1R. Since we trust the productive test more (as hinted at in H1), the non-significant group score on the guess-prone T1R appears less damaging to the hypotheses explored here.

The expectation that participants would perform better with NCMs of HF than LF nouns (H2(b)) was confirmed on both tasks. This is hardly surprising since H2(a) already indicated relatively poor performance on NCMs of LF+NCDs, with participants preferring alternative solutions, even at the risk of ungrammaticality or noncompliance with instructions. Since the results of our original hypothesis concerning NCMs of HF+NCDs showed that their NCMs were very well-known, a different outcome on H2(b) would have been unlikely.

Both tasks showed a HF advantage, contrary to our "equal-performance" expectation (H2(c)). Participants scored better on CMs of HF+CDs than subordinate CMs of LF+NCDs. We do not think this was caused by NCM dominance over CMs in the LF nouns. But we do think that the confidence needed to apply constructive skills to form LF targets cannot measure up to that needed to recruit frequently-experienced targets from memory. Except in H1(b), all the frequency-sensitive comparisons of CMs went in favour of HF items. Examples like *snoozefest, crapper*$_{ag}$ etc. from T2P suggest that learners seek solutions they are more experienced with, even if this means breaking a few rules.

6. General discussion and conclusion

The present study set out to test some UBM assumptions about suffix acquisition by EFL learners. Our results indicate:

- CM advantage (H1a: T1R and T2P, H2a: T2P), and
- HF advantage (H1b on T2P; H1c, H2b and H2c on both tasks).

Effect sizes, ranging from large to huge for all statistically significant results demonstrate that both HF and CM are important factors in EFL acquisition of *-er.* Thus, our general findings support the UBM predictions stated at the beginning.

CM advantage, revealed by all but one relative group comparisons (exception is H2a on T1R) potentially indicates participants' command of CM schemas, i.e. their sensitivity to the high type frequency of agent and instrument *-ers*. However, good performance on CMs is also compatible with lexical knowledge if factors other than objective frequency in COCA are allowed (see below).

HF advantage, revealed by most independent group comparisons of LFs with their HF standards on both tasks indicates participants' sensitivity to token frequencies. This supports earlier findings of frequency effects on lexical production and recognition in both L1 and FL contexts (Ellis, 2002a and references cited therein; Kirsner, 1994).

However, for all their merits, group results under-represent the complexity of learning morphology in EFL. They hide non-compliant detail that not only suggests the complexity of usage-based EFL learning, but also foregrounds other relevant factors, unrelated to input.

First, if learners do command CM schemas, in theory one could expect 100% performance on all LF agents/instruments since schemas would compensate for missing lexical knowledge. However, cases like *laugher*$_{ag}$, *refresher*$_{ag}$, *crapper*$_{ag}$ on T2P suggest that schema-based knowledge, if available, easily buckles to strongly-entrenched, sometimes not even fully compliant, lexical competitors (on token blocking as a constraint on productivity see Aronoff, 1976; Plag, 1999; Rainer, 2005).

Second, NCM overperformers, especially those on which participants did well on both tasks (*weeper*$_{caus}$, *yawner*$_{caus}$, *crapper*$_{loc}$) suggest that some items may have subjective salience for our student-participants. Thus, no matter how useful as a usage-based operationalization of grammar, objective input frequencies do not explain everything that makes forms worthy of intake. Whether a more targeted, youth-oriented corpus would better reflect our findings should be explored in future.

Third, although we did not aim to explore task-induced variation, we found differences between production and recognition. Generally, participants' performance on NCMs dropped and that on CMs increased in T2P. We attributed this to the lesser susceptibility of T2P to abrupt decision-making characteristic of T1R. On the yes-no task participants had a 50:50 chance of a correct guess with NCMs. Whether there are psychological factors at play here – tendency to accept the new, 'atypical'—must be left for future research. One thing is certain, participants did fare worse with NCMs when they had to produce them. Building a complex form requires putting more thought into the process and navigating one's way through a number of possibilities (verbs can take different suffixes or be converted, participants even used syntactic paraphrase despite instruction to supply one word only). This is where guesswork is less likely to lead to success.

Finally, let us not forget that variability, as we found with CM, but especially NCMs, is typical in EFL and cannot be excluded despite the relative homogeneity of participants (R. Ellis, 1994). Unlike L1 development, EFL is characterized by unique, non-linear learning paths and, particularly, incompleteness. This is attributed to numerous cognitive and affective factors, as well as language instructional (i.e. explicit vs. implicit training) and exposure conditions, all of which interact and shape progress and ultimate attainment (Tagarelli et al., 2016). Their precise role, and the role of additional factors like task-induced variation, in the acquisition of *-er* (forms), i.e. how they interact with each other and with input frequencies remains to be explored in future.

What can we conclude about our advanced students' knowledge of *-er*? We did get results suggesting consistently good performance on the CMs of nouns that our participants had possibly not heard before. But since we do not have access to participants' usage/learning histories, we cannot be sure about the source of this knowledge: whether the CM effects observed are (only) due to schematic knowledge of the suffix or (also) lexical knowledge of at least some of these nouns. Additional experiments using nonce-words could shed more light on this question. As for NCMs, even allowing for the very real chance of guesswork on T1R, our results suggest that participants do have isolated knowledge of some nouns but no ability to easily and consistently generalize to new instances, i.e. no NCM schemas. Whether explicit instruction could promote the learning of 'new' associations of *-er* with NCMs (Ellis, 2002a) should also be verified in future, if not for our learners' proficiency gain, then at least for the sake of theoretical discussion.

With this study we hope to have contributed to a better understanding of some aspects of derivational proficiency in advanced EFL learners. We have confirmed that frequency is one important factor. However, the study also indicated that frequency does not explain everything, thus reaffirming views that a myriad of linguistic and non-linguistic factors should be integrated into models of FL learners' grammatical proficiency (Ellis, 2012).

References

Adams, Valerie (2001). *Complex Words in English*. Harlow, Essex: Longman.

Anić, Vladimir (2004). *Veliki rječnik hrvatskoga jezika*. Zagreb: Novi Liber.

Aronoff, Mark (1976). *Word Formation in Generative Grammar*. Cambridge, Massachusetts / London, England: The MIT Press.

Baayen, Harald, Rochelle Lieber (1991). Productivity and English derivation: a corpus-based study. *Linguistics* 29: 801–844. https://doi.org/10.1515/ling.1991.29.5.801

Babić, Stjepan (2002). *Tvorba riječi u hrvatskome književnome jeziku,* 3rd improved edition. Zagreb: HAZU Nakladni zavod Globus.

Barlow, Michael, Suzanne Kemmer (eds.). (2000). *Usage-Based Models of Language.* Stanford: CSLI.

Bley-Vroman, Robert (2002). Frequency in production, comprehension and acquisition. *Studies in Second Language Acquisition* 24: 209–213. https://doi.org/10.1017/S027226310200205X

Brysbaert, Marc, Boris New (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41: 977–990. https://doi.org/10.3758/BRM.41.4.977

Bybee, Joan, Carol Lynn Moder (1983). Morphological classes as natural categories 59(2): 251–270.

Bybee, Joan (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* 10:5: 425–455. https://doi.org/10.1080/01690969508407111

Bybee, Joan (2007). *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195301571.001.0001

Bybee, Joan (2008). Usage-based grammar and second language acquisition. Robinson, Peter, Nick Ellis, eds. *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York: Routledge, 216–236.

Bybee, Joan (2010). *Language, Usage and Cognition.* Cambridge: Cambridge University Press. https://doi.org/10.1111/j.1540-4781.2012.01338.x

Clark, Eve V. (2014). Acquisition of Derivational Morphology. Lieber, Rochelle, Pavol Štekauer, eds. *The Oxford Handbook of Derivational Morphology*. Oxford: Oxford University Press, 424–439. doi:10.1093/oxfordhb/9780199641642.013.0024

Clark, Eve V., Sophia R. Cohen (1984). Productivity and memory for newly formed words. *Journal of Child Language* 11: 611–625. https://doi.org/10.1017/S0305000900005985

Cohen, Jacob (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cruse, Alan, William Croft (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Davies, Mark (2008–). *The Corpus of Contemporary American English: 450 million words*, 1990-present. Available online at http://corpus.byu.edu/coca/.

Divjak Dagmar, Catherine L. Caldwell-Harris (2015). Frequency and entrenchment. Dabrowska, Eva, Dagmar Divjak, eds. *Handbook of Cognitive Linguistics*. Berlin: de Gruyter, 53–75.

Dörnyei, Zoltán, Stephen Ryan (2015). *The Psychology of the Language Learner Revisited*. New York/London: Routledge. https://doi.org/10.4324/9781315779553

Ellis, Nick C. (2002a). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition* 24: 297–339. https://doi.org/10.1017/S0272263102002140

Ellis, Nick C. (2002b). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24: 143–188. https://doi.org/10.1017/S0272263102002024

Ellis, Nick C. (2008). Usage-based and form-focused language acquisition. Robinson Peter, Nick C. Ellis, eds. *Handbook of Cognitive Linguistics and Second Language Acquisition*. NewYork: Routledge, 372–405.

Ellis, Nick C. (2012). Frequency-based accounts of SLA. Gass, Susan M., Alison Mackey, eds. *Handbook of Second Language Acquisition*. London/New York: Routledge, 193–210.

Ellis, Nick C., Richard Schmidt (1997). Morphology and longer distance dependencies: laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition* 19: 145–171.

Ellis, Rod (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Fillmore, Charles (1968). The case for case. Bach, Emmon, Robert Harms, eds. *Universals in Linguistic Theory*. New York: Holt, Rinehard and Winston, 1–88.

Gentner, Dedre, Arthur B. Markman. (1997). Structure mapping in analogy and similarity. *American Psychologist* 52(1): 45–56. https://doi.org/10.1037/0003-066X.52.1.45

Gentner, Dedre, José Medina (1998). Similarity and the development of rules. *Cognition* 65(2-3): 263–297. https://doi.org/10.1016/S0010-0277(98)00002-X

Glynn, Dylan (2014). Polysemy and synonymy: Cognitive theory and corpus method. Glynn, Dylan, Justina A. Robinson, eds. *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*. Amsterdam/Philadelphia: John Benjamins, 7–38.

Gries, Stefan Th. (2008). Corpus-based methods in analyses of Second Language Acquisition data. Robinson, Peter, Nick C. Ellis, eds. *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York: Routledge, 406–431.

Hay, Jennifer (2001). Lexical frequency in morphology: is everything relative? *Linguistics* 39: 1041–1070.

Kirsner, Kim (1994). Implicit processes in second language learning. Ellis, Nick, ed., *Implicit and Explicit Learning of Languages*. San Diego, CA: Academic Press, 283–312.

Kline, Rex (2010). *Principles and Practice of Structural Equation Modeling*. (3rd ed.). New York, London: The Guilford Press.

Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar*, Vol. I. Stanford: Stanford University Press.

Langacker, Ronald W. (2000). A dynamic usage-based model. Barlow, Michael, Suzanne Kemmer, eds. *Usage-Based Models of Language*. Stanford: CLSI Publications, 1–64.

Lieber, Rochelle (2004). *Morphology and Lexical Semantics.* Cambridge: Cambridge University Press.

Lieber, Rochelle (2014). Methodological issues in studying derivation. Lieber, Rochelle, Pavol Štekauer, eds. *The Oxford Handbook of Derivational Morphology*. Oxford: Oxford University Press, 84–94. https://doi.org/10.1093/oxfordhb/9780199641642.013.0006

Lowie, Wander (2000). Cross-linguistic influence on morphology in the bilingual mental lexicon. *Studia Linguistica* 54: 175–185. https://doi.org/10.1111/1467-9582.00058

Lowie, Wander (2005). Exploring a second language. The discovery of morphological productivity. *EUROSLA Yearbook 5*: 251–268. https://doi.org/10.1075/eurosla.5.12low

Lowie, Wander, Marjolijn Verspoor (2004). Input versus transfer? The role of frequency and similarity in the acquisition of L2 prepositions. Achard, Michel, Susanne Niemeier, eds. *Cognitive Linguistics, Second Language Acquisition, and Foreign Language Teaching*. Berlin: Mouton de Gruyter, 77–93. https://doi.org/10.1515/9783110199857.77

Luschützky, Hans Christian, Franz Rainer (2013). Instrument and place nouns: A typological and diachronic perspective. *Linguistics* 51(6): 1301–1359. https://doi.org/10.1515/ling-2013-0051

MacWhinney, Brian (2008). A unified model. Robinson, Peter, Nick C. Ellis, eds. *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York: Routledge, 341–371.

Marchand, Hans (1969). *The Categories and Types of Present-Day English Word-formation*. München: C. H. Beck'sche Verlagsbuchhandlung.

Morris, Scott, Richard P. DeShon (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods* 7: 105–125. https://doi.org/10.1037//1082-989X.7.1.105

Nagy, William, Irene-Anna Diakidoy, Richard Anderson (1993). The acquisition of morphology: Learning the contribution of suffixes to the meanings of derivatives. *Journal of Reading Behavior* 25: 155–170.

Panther, Klaus U., Linda U. Thornburg (2002). The roles of metaphor and metonymy in English *-er* nominals. Dirven, René, Ralf Pörings, eds. *Metaphor and Metonymy in Comparison and Contrast*. Berlin/New York: Mouton de Gruyter, 279–319.

Plag, Ingo (1999). *Morphological Productivity. Structural Constraints on English Derivation*. Berlin/New York: Mouton de Gruyter.

Rainer, Franz (2005). Constraints on productivity. Štekauer, Pavol, Rochelle Lieber, eds. *Handbook of word-formation*. Spinger Verlag, 335–352.

Robinson, Peter, Nick Ellis, eds. (2008). *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York: Routledge.

Ryder, Mary Ellen (1991). Mixers, mufflers and mousers: The extending of the -er suffix as a case of prototype reanalysis. *Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society*. Berkeley, California: Berkeley Linguistics Society, 299–311. https://doi.org/10.3765/bls.v17i0.1616

Sawilowsky, Shlomo S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods* 8(2): 597–599.

Silić, Josip, Ivo Pranjković (2007). *Gramatika hrvatskoga jezika za gimnazije i visoka učilišta*. (2nd edition). Zagreb: Školska knjiga.

Tagarelli, Kaitlyn M., Simon Ruiz, José Luis Moreno Vega, Patrick Rebuschat (2016). Variability in second language learning: The roles of individual differences, learning conditions, and linguistic complexity. *Studies in Second Language Acquisition* 38: 293–316. http://dx.doi.org/10.1017/S0272263116000036

Taylor, John R. (2002). *Cognitive Grammar*. New York: Oxford University Press.

Talmy, Leonard (2000a). *Toward Cognitive Semantics, Volume 1: Concept Structuring Systems*. Cambridge, MA: MIT Press

Talmy, Leonard (2000b). *Toward Cognitive Semantics*, Volume 2: *Typology and Process in Concept Structuring*. Cambridge, MA: MIT Press.

Tomasello, Michael (2000). First steps in a usage-based theory of language acquisition. *Cognitive Linguistics* 11: 61–82. https://doi.org/10.1515/cogl.2001.012

Tomasello, Michael (2003). *Constructing a Language. A Usage-Based Theory of Language Acquisition*. Cambridge, MA/London, UK: Harvard University Press.

Tyler, Andrea (2010). Usage-based approaches to language and their applications to second language learning. *Annual Review of Applied Linguistics* 30: 270–291. https://dx.doi.org/10.1017/S0267190510000140

Van Valin, Robert, David Wilkins (1996). The case for 'effector': Case roles, agents and agency revisited. Shibatani, Masayoshi, Sandra A. Thompson, eds. *Grammatical Constructions*. Oxford: Oxford University Press, 289-322.

Yoonsang Song, Do Younngah, Arthur L. Thomspon, Eileen R. Waegemaekers, Jongbong Lee (2019). Second language users exhibit shallow morphological processing. *Studies in Second Language Acquisition* 2020: 1–16. https://doi.org/10.1017/S0272263120000170

COCA word frequency list. Retrieved from http://www.wordfrequency.info/free.asp?s=y

Authors' address:

Višnja Pavičić Takač,
University of Osijek,
Faculty of Humanities and Social Sciences
Osijek, L. Jägera 9, 31 000 Osijek, Croatia;
vpavicic@ffos.hr

Gabrijela Buljan,
University of Osijek,
Faculty of Humanities and Social Sciences
Osijek, L. Jägera 9, 31 000 Osijek, Croatia;
gbuljan@ffos.hr

Received: March 16, 2023

Accepted for publication: April 28, 2023

Appendix A

Type frequencies of five semantic categories

		SR1	SR2	SR3	SR4	TOTALS	%
central meaning	**agent**	955	212	8	-	1175	63.8
	instrument	335	227	11	-	573	31.11
noncentral meaning	**patient**	12	25	20	1	58	3.15
	location	7	5	4	-	16	0.86
	cause	4	8	2	1	15	0.81
	other	2	3	-		5	0.27
	TOTALS	1315	480	45	2	1842	

SR = semantic role

Appendix B

Participants (selected descriptive statistics) (N = 127)

			Min	*Max*	*Mean*	*SD*
age			19	29	21.05	1.72
classroom contribution to EFL knowledge (%)			4	100	42.38	18.37
authentic input contribution to EFL knowledge (%)			0	95	56.98	18.67
motivation (1 – not motivated to 4 – very motivated)			1	4	3.43	.61
exposure to authentic English	reading English net	days per week	1	7	6.05	1.50
		hours per day	1	6	3.65	1.61
	reading magazines	days per week	0	7	3.70	2.55
		hours per day	0	6	1.87	1.38
	reading literature	days per week	0	7	2.97	2.058
		hours per day	0	6	2.50	1.11
	listening/watching	days per week	1	7	6.38	1.11
		hours per day	1	6	3.79	1.61