Doctoral (PhD) dissertation

Approaches of cognitive linguistics and ontologies in lexicographic sense delineation

Esra Moustafa Abdelzaher Abdelkareem

Supervisor: Dr. Ágoston Tóth

University of Debrecen

Institute of English and American Studies

Debrecen, Hungary

2023

Approaches of cognitive linguistics and ontologies in lexicographic sense delineation

Dissertation submitted in partial fulfilment of the requirements for the doctoral (PhD) degree in **linguistics**

Written by **Esra Moustafa Abdelzaher Abdelkareem** certified **holder of a Master's degree in English linguistics**

Prepared in the framework of the **Doctoral School of Linguistics** of the University of Debrecen (**Doctoral Programme of General and Applied Linguistics, Subprogramme in English Linguistics**)

Dissertation advisor: **Dr.Tóth Ágoston**

The official opponents of the dissertation:

Dr. ……..
Dr. ……..
Dr.…… ..

The evaluation committee:

Chairperson: Dr. ……..
members: Dr. ……..
Dr. ……..
Dr. ……..
Dr. ……..

The date of the dissertation defence: …20 ……

Declaration

"The undersigned Esra’ Moustafa Abdelzaher Abdelkareem, in full awareness of my liability, I declare that the dissertation submitted is my original academic work, which was prepared with attention to the international norms of copyright, and the references included in it are clear and complete. I further declare that I am not currently subject to a procedure aimed at the revocation of a doctoral degree, and that a doctoral degree already awarded has not been revoked from me in the past 5 years. I have not submitted the present dissertation at another institute before, and it has not been rejected”

Acknowledgments

I would like to express my sincere gratitude to my supervisor, **Dr. Tóth Ágoston**, for his continuous help and support during the past five years. His genuine effort to concretize my ideas, work out my research plans, and fulfill my academic objectives is invaluable. As a mentor, not only a supervisor**, Dr. Tóth Ágoston's** guidance was never confined to the PhD dissertation. He encouraged me to engage in multiple research activities and stay updated on the usability of technological advances in language-related investigations. I learned innumerable things from **Dr. Tóth Ágoston**, including probing theoretical and practical solutions for research challenges, exploring linguistic phenomena from various perspectives, and being open to all suggestions.

I would also like to extend my thanks to **Prof. Dr. András Kertész**, **Dr. Péter Pelyvás, Dr. Katalin P. Márkus**, **Dr. Tibor M. Pintér, Dr. Attila Cserép and Dr. Erzsébet Balogh** for spending time and effort reviewing the earlier draft of this dissertation and providing their sage advice and recommendation for possible improvements. Their commitment to excellence in academic research has been a constant source of motivation.

Contents

List of Abbreviations

Abbreviation	Stands for
AHD	American Heritage Dictionary of the English Language
BAWE	British Academic Written English
BNC	British National Corpus
CEFR	Common European Framework of Reference for Languages
COCA	Corpus of Contemporary American English
COD	Concise Oxford Dictionary
ESL	English as Second Language
Fes	Frame Elements
FN	FrameNet
GDEX	Good Dictionary Example
ISD	Idiomatic and Syntactic Dictionary
KWIC	Key Words In Context
LDOCE	Longman Dictionary of Contemporary English
LOB	Lancaster-Oslo-Bergen corpus
LUs	Lexical Units
MECD	Macmillan English-Chinese Dictionary for Advanced Learners
MIP	Metaphor Identification Procedure
MLD	Monolingual Learners' Dictionary
MWEs	Multiple Word Expressions
NLP	Natural Language Processing
NODE	New Oxford Dictionary of English
OALD	Oxford Advanced Learners' Dictionary
OALECD	Oxford Advanced Learner's English-Chinese Dictionary
OED	Oxford English Dictionary
Synset	Synonymy set
TEFL	Teaching English as Foreign Language
WN	WordNet
WSD	Word Sense Disambiguation

Chapter One – Introduction

Linguists present several arguments regarding what constitutes word meaning. Lyons (1977) pointed to several theories of word meaning. For instance, the denotational theory[1] associates the meaning of an expression with what the expression denotes or stands for in the real world. In contrast, the ideational theory[2] equates the meaning of an item with the idea connected with it in the mind of language users. Also, the meaning-in-use theory[3] argues that the use of an expression in a context identifies, or is equivalent to, the meaning of this expression.

Over time, linguists and lexicographers started to adopt more conflicting views on meaning to the extent that some rejected the possibility of having a word meaning outside the context of use. In the traditional view, words are believed to have several types of meanings, such as lexical and contextual meanings. Lexical meaning is the semantic content of the word regardless of the context in which it may be used. In contrast, contextual meaning arises when the word is used in real communicative situations. Bergenholtz and Gouws (2017) revealed how linguists such as Louw reasoned that only the lexical meaning of a word should be listed in dictionaries, whereas the interpretations that emerge with relevance to the use of a word in a context should not be part of its meaning. Accordingly, lexical meaning and the so-called contextual meaning are not two types of meaning. They rather represent word meaning (as recorded in dictionaries) and word use. The authors contrasted the contradictory views of Louw and Hanks on contextual meaning. On the one hand, Louw did not consider contextual meaning a part of the meaning of the word. On the other hand, Hanks argued against the existence of an abstract word meaning outside the context of use. That is to say, meaning is not a property already inherent in a word. It is an interpretation a word acquires in a context.

Regardless of the theoretical disagreement over the (non)existence of word meanings, the lexicographer has to enumerate the senses of words in dictionary entries. Sense enumeration is both a tradition and a practical necessity. In this regard, lexicographers have to find a way to identify what constitutes a word sense. Also, they have to decide whether the differences in word

[1] The denotational theory of meaning has its roots in the Fregean logic.

[2] The ideational theory of meaning has been mainly developed by the philosopher John Locke.

[3] The meaning-in-context theory is largely dependent on the Gricean pragmatics.

uses are distinct enough to be recorded as individual word senses (i.e., split from other uses) or are minor and should be lumped with another meaning (Tóth, 2006).

1.1 Statement of the Problem

Lexicography has a long history of defining meaning and has witnessed considerable developments in the past 50 years. Nevertheless, lexicographers still have challenging tasks to deal with and dictionary gaps to address.

1.1.1 Challenging Tasks in Lexicography

Kilgarriff (1998) identified the most challenging tasks lexicographers faced during their work on the third edition of LDOCE. In his short report, he considered the most complicated tasks to be the ones that lack clear rules or guidelines. He categorized the tasks into (a) analysis, i.e., tasks relevant to the pre-writing stage and the analysis of word behavior in context, and (b) synthesis, i.e., tasks relevant to the content that will finally be presented in the dictionary. The most challenging task during the analysis stage was "splitting; identifying senses of a word". Dictionary style guides contain no instructions on splitting or lumping senses in an entry (Atkins and Rundell, 2008). At that time (1998), there was no helpful information on how to deal with this problem in books. Therefore, lexicographers used to, and perhaps still, depend on their experience. Krishnamurthy (1996) listed the 32 papers that detailed the policy of making the COBUILD dictionary, and none of them dealt with the sense delineation problem. Accordingly, the task ranked second in the list of the most challenging tasks after formulating definitions.

Kilgarriff (2007) further pointed out the problems lexicographers encounter when identifying senses. He chose the BNC because it was created to be a balanced representation of contemporary English. Accordingly, the most frequent uses in this corpus should be reflected in the senses of the word in a general-purpose dictionary. However, difficulties appear when they attempt to abstract the senses from corpus citations. They encounter metaphors at different levels (e.g., words, sentences, or discourse). Using the target word as part of a name or sublanguage is likewise problematic for lexicographers. Lexicographers have to decide whether this is a different unpredictable sense that should be recorded in a dictionary or not. Moreover, non-standard word use always depends on deviation from the known use. However, the new use is not always salient for users, specifically if triggered by a combination of words rather than a single target word.

1.1.2 Gaps in Dictionaries

The most challenging problems detected by Kilgarriff (i.e., wording and identifying senses) are also reflected in the most reported dictionary gaps. Hanks (2013), for example, criticized definitions in Merriam-Webster's *New International Dictionary* for three drawbacks common in modern lexicography. When wording definitions, lexicographers may include the accidental properties of a word because they confuse them with the essential properties. The criticized dictionary included both "sweet" and "used as a condiment and preservative for other foods and for drugs" in the definition of *sugar*. This wording treats the essential properties of *sugar*, which are part of its definition like the accidental properties, which are irrelevant to the definition. Second, some definitions prioritize scientific accuracy over the accessibility of the definition to the average reader. Third, definitions also reflect lexicographers' disregard of the theoretical advances relevant to the fuzziness and variability of word meaning.

The problem of identifying senses is likewise mirrored in dictionary entries. The problem becomes salient when dictionaries are used for NLP tasks. Current definitions of word senses in dictionaries do not reflect the dynamicity and flexibility of the natural language. The list of senses provided for a word misleadingly suggests the finiteness of word meanings. Describing word senses should show a word's dynamicity and meaning potential instead of listing a finite set of definitions that do not correspond to the linguistic reality. In addition, the various definitions of word senses do not provide any clues as to how one sense is distinguished from another. Complicating matters, the definitions are not mutually exclusive, and they usually overlap (Hanks, 2019).

Moreover, attempts at matching actual word uses in a corpus to dictionary senses usually retrieve unsatisfactory results. Such attempts display the mismatches between the linguistic reality and the senses in dictionaries. Mapping a sample of words, occurring from 26 to 29 times in the Lancaster-Oslo-Bergen (LOB) corpus, to their senses in the LDOCE revealed that at least one usage of 87% of the words did not match any of the dictionary senses. In addition, a single citation could be mapped to more than one dictionary sense, and it was rare for a citation to clearly match a single sense (Kilgarriff, 1992). It is worth mentioning that LOB is a one-million corpus of British English that covers 15 genres. Therefore, it was supposed to be a representative sample of the British variety.

The same gap was detected by Palmer, Ng and Dang (2007), who acknowledged inevitable shortcomings in all inventories (e.g., dictionaries) if exploited in WSD tasks. Natural word uses are often implied in underspecified contexts; accordingly, multiple senses can be applied to them. Therefore, the authors suggested providing both human and automatic annotators with two more options rather than selecting a single sense for each word use. They recommended the acceptability of choosing more than one sense or a broader sense that embeds several meaning specifics.

Dictionary senses do not accurately represent the ambiguity faced by language users in natural contexts. Dictionaries include fine-grained distinctions a language user may not be fully aware of, and they list rare senses a language user may not be familiar with (Britton, 1978). In addition, many fine-grained senses of polysemous words cause information overload and perplex inexperienced dictionary users (Gouws and Tarp, 2017). In general, the attempts of dictionary makers to stimulate the mental lexicon of a native speaker are not successfully implemented in the dictionary senses that do not perfectly reflect language reality (Lew, 2013).

1.2 Objectives

The current study explores the different approaches to the challenge of sense delineation in a Monolingual Learners' Dictionary (MLD), in our case, the Oxford Advanced Learners' Dictionary (OALD), and two lexical databases, which are FrameNet (FN) and WordNet (WN). OALD represents the lexicographic baseline in this study. In contrast, FN and WN are included to test the possible contributions of cognitive and lexical semantics to traditional lexicographic practice (especially the challenge of sense delineation). The study aims at reaching the following objectives:

(1) Investigating the effectiveness of using the cognitive semantic approach proposed by Frame Semantics and implemented in FN in meeting the challenge of sense delineation in lexicography;

(2) Exploring the effectiveness of using the lexical semantic information in WN for delineating senses in lexicography;

(3) Examining the applicability of integrating lexicographic information from OALD, cognitive semantic information from FN and sense relations from WN in the same lexicographic entry; and

(4) Studying the usefulness of examples to the identification of word senses in OALD, FN and WN.

This dissertation discusses meaning identification from lexicographic, cognitive and ontological perspectives. After this introductory chapter, chapter two focuses on meaning as recorded in dictionaries. It addresses the uses, intended users, typology and structures of dictionaries. Chapter three discusses the contributions of lexical semantic information to meaning identification in lexical databases and ontologies. The WordNet database (WN) will be addressed in more detail because it is the most used linguistic ontology in lexicographic research. Chapter four focuses on the contributions of cognitive linguistics to sense delineation. It primarily focuses on the Frame Semantics approach to sense delineation because it has already been implemented in a database that can be used for lexicographic purposes, i.e., FrameNet (FN).

After the theoretical discussions, chapter five presents a classroom-based experiment designed to practically test the effectiveness of OALD, FN and WN in a dictionary-consultation context. Chapter six reports a second experiment demonstrating the results of using a hybrid dictionary entry (based on OALD, FN and WN) by ESL university students. Chapter seven describes a classroom experiment that was designed to test the usefulness of example sentences in OALD, FN and WN in illustrating the delineated senses. Statistical tests (mainly Pearson correlation, one-way ANOVA and Post Hoc Tukey tests) are used to validate or reject the hypotheses, and Cronbach's Alpha was used to measure the reliability of each question.

Chapter Two – Dictionaries and Word Meaning

A dictionary is a specific type of book that is typically consulted to know or double-check information about words (Bejoint, 1993). A dictionary is basically used for encoding or decoding reasons. In a typical decoding context, the user consults a dictionary to understand the meaning of a word or translate a text into his or her first language. The encoding context is relevant to the correct use of a word or the translation of a text into a foreign language. This binary use is frequently referred to as comprehension and production tasks (Atkins and Rundell, 2008).

Bergenholtz and Agerbo (2018) proposed another classification of dictionary uses in their dictionary typology. They define lexicographic resources such as dictionaries as utility tools that meet specific user needs. A lexicographic tool such as a dictionary or an encyclopedia is used to satisfy operative, comprehension or knowledge needs arising from a variety of para-lexicographical situations. Para-lexicographical situations range from reading traffic signs to translating a text in an academic or professional context. The operative use of a dictionary is performing a translation task. Comprehension needs correspond to consultation for decoding purposes (i.e., understanding the meaning of a word). Knowledge-related needs include expanding the user's knowledge of a particular topic through encyclopedias or specialised dictionaries. The suggested taxonomy still reflects the decoding use in "comprehension" and the encoding use in "operation". However, it adds "knowledge" as a different reason for using dictionaries. Although knowledge-related uses do not directly correspond to the encoding and decoding uses mentioned by Atkins and Rundell (2008), the closest match in the traditional proposal may be "studying a particular subject".

2.1 Dictionary Typology

Dictionaries are primarily commercial or scholarly. Most lexicographic research is directed to commercial or trade dictionaries that target public users. Scholarly or historical dictionaries, in contrast, are not intended for decoding or encoding purposes. They are a "cultural index" as Hanks (2013) described them. The criteria for categorizing dictionaries are usually based on and applicable to trade dictionaries.

Atkins and Rundell (2008) proposed a dictionary typology based on eight categories. Each category consists of a set of properties that are not mutually exclusive but applicable to all

dictionaries. That is to say, a dictionary essentially has at least one property from each category. The language of a dictionary classifies it as monolingual, bilingual, bilingualized or multilingual. The coverage of a dictionary categorizes it as a dictionary of idioms or legal terms, for instance. In addition, the dictionary's size (e.g., standard or concise) and medium (e.g., print, electronic, web-based) must be taken into consideration. However, nowadays, print and CD dictionaries are not as commonly used as they were at the time of writing the book in 2008. The organization of information in a dictionary is another factor that groups a dictionary in the standard category that uses word-to-meaning organization or the less familiar set using word-to-meaning-to-word ordering. Other factors are related to the target user's language (e.g., native or non-native English speakers) and skills (e.g., learners or professionals). Finally, a dictionary is classified based on its purpose, i.e., decoding, encoding or both.

Tarp (2018b) limited the criteria for defining and classifying a dictionary to "form, content and purpose". "Form" includes the traditional criteria of medium (e.g., print book or digital dictionary) and language (e.g., monolingual) as well as the organization of the wordlist (e.g., alphabetical). "Content" classifies dictionaries into language and encyclopedic types. Finally, "purpose" defines four functions a dictionary might perform. First, communicative functions cover any written or oral communication that needs a dictionary, such as translation, text understanding or production. Communicative functions combine encoding and decoding uses. Second, cognitive functions are generally defined as functions of disseminating knowledge among users. Third, operative functions help users perform specific tasks. Fourth, interpretive functions help users understand non-linguistic signs.

A more recent typology is proposed by Bergenholtz and Agerbo (2018). They broaden their typology to include any lexicographic tool, i.e., "a utility tool" devised for meeting user needs such as operative, comprehension or knowledge-expansion purposes. Therefore, the taxonomy puts translation dictionaries and spelling checkers in the same hierarchical node (i.e., a lexicographic tool designed to help users with a specific operative task). In contrast, Wikipedia, for instance, is placed in another node (a lexicographic tool designed to help users gain knowledge about a certain topic). In this context, a lexicographic tool comprises three types, based on its function in this context. A tool can be (a) poly-functional to serve multiple needs, (b) mono-functional to address a single need or (c) individual —a tool designed by users to tackle their consultation needs in certain contexts.

After reviewing several dictionary taxonomies (based on formal and functional characteristics), Bejoint (1993) stated that dictionaries could not be classified according to a single taxonomy. He was still able to make binary distinctions between dictionaries with reference to the microstructure and the macrostructure. The distinctions, first, included general and specialized dictionaries. A general macrostructure contains a relatively comprehensive list of words that are not restricted to a certain variety, topic or time. A general microstructure provides definitional, grammatical, collocational, phonological and etymological information, among others, in all entries. Specialized dictionaries specify their wordlists to a certain dialect, topic or linguistic phenomenon and restrict the information in the entries to certain types. Second, monolingual dictionaries use a single language in the wordlist and use the same language to provide information at the microstructural level. Bilingual dictionaries make use of two languages generally in the form of source-target equivalents (i.e., translation dictionaries). Third, encyclopedic dictionaries include, at the macrostructural level, names of countries and people as word entries, unlike general language dictionaries. He also distinguished learners' dictionaries from native ones and adult dictionaries from the ones directed to children.

Landau (1993) also surveyed the different types of dictionaries and proposed several criteria to categorize dictionaries, such as the number of the languages in the dictionary, the age of the target users, the scope of the covered topics, the size of the included lexical units and the linguistic approach of the dictionary (i.e., descriptive or prescriptive).

For the purposes of this dissertation, MLDs, new Internet-based dictionaries, and lexical databases will be discussed in more detail in the following sections.

2.1.1 Monolingual Learners' Dictionaries

Although the purpose of compiling a dictionary is not always identical to the reasons for using it, the two aspects are interrelated. Hanks (2013) introduced a binary classification of the reasons for compiling dictionaries in the history of lexicography. Compilers either aimed at providing a "cultural index" of their languages (e.g., *Oxford English Dictionary*, OED hereafter) or helping learners of a foreign language (e.g. *Idiomatic and Syntactic Dictionary*, ISD hereafter). Language teaching has always been a motivation for compiling dictionaries and, accordingly, a purpose for using them.

The history of compiling MLDs started with the work of Hornby, Palmer and West, who were prominent English teachers. Whereas Hornby and Palmer were teaching English to Japanese learners in Tokyo, West was based in India. West has been interested in defining the core vocabulary of English using a limited number of words. He was a pioneer in the limited vocabulary control movement. His dictionary, the *New Method English Dictionary*, used 1490 words to define 24000 headwords. This policy was followed later by most MLDs publishers, who usually depend on a set of 2000-3000 words to define the headwords. The first edition of *Cambridge International Dictionary of English,* for instance, used around 2000 defining words. However, the second most known dictionary for following the vocabulary control movement was the *Longman Dictionary of Contemporary English* (LDOCE), 1st edition in 1978 (Miller, 2018). To date, the online version of the LDOCE sticks to the use of simple defining words if compared to other learners' dictionaries. The following are the definitions of *inference*.n in the online versions of LDOCE, Collins COBUILD and OALD. Whereas Collins COBUILD and OALD use a B1-level word in the definition (*conclusion* and *indirectly*), LDOCE strictly uses A1 words in the definition of the same word, i.e., *inference*. According to Oxford 3000 worlist of the core vocabulary for English learners, the LDOCE definition uses more basic vocabulary than COBUILD and OALD.

Table 1. The definitions of *inference* in the online versions of LDOCE[4], Collins COBUILD[5] and OALD[6]

LDOCE	*Inference*: 1. something that you think is true, based on information that you have
Collins	*Inference:* An inference is a conclusion that you draw about something by using information that you already have about it.
OALD	*Inference:* something that you can find out indirectly from what you already know

Hornby and Palmer were more concerned with adding grammatical information to their dictionaries. Although Hornby intuitively used simple defining vocabulary and avoided rare and difficult words in the definitions, he is best known for his inclusion of syntactic information in dictionaries. In 1942, Hornby was in contact with Palmer, who led research on English collocations and verb complementation at the Tokyo Institute for Research in English Teaching.

[4] The dictionary is available at https://www.ldoceonline.com/ (accessed on 15-1-2023).

[5] The dictionary is available at https://www.collinsdictionary.com/ (accessed on 15-1-2023).

[6] The dictionary is available at https://www.oxfordlearnersdictionaries.com/ (accessed on 15-1-2023).

The results of this research appeared in Palmer's dictionary, *A Grammar of English Words*, and Hornby's *Idiomatic and Syntactic Dictionary*. Both dictionaries provided valuable syntactic information about the complementation of English words. They aimed at helping L2 learners with productive tasks (Jackson, 2003). It was Hornby's dictionary that made a significant contribution to the development of MLDs because Oxford undertook the mission of republishing it in 1948, 1952 and most notably in 1947 under the title *Oxford Advanced Learner's Dictionary.*

After the great success of the orientation towards L2 learners, MLDs started to focus more on providing grammatical information to intermediate and advanced learners. In 1978, the first edition of the *Longman Dictionary of Contemporary English* was introduced. It followed the principles of West, Hornby and Palmer. The dictionary makers stated that only 2000 words were used to define all the headwords in the dictionary. Grammatical information was provided systematically using simple codes such as "T" for transitive. Usage notes on currency, frequency, and sociolinguistic variations were also one of the innovative user-oriented features in the dictionary (Heuberger, 2018).

In 1987, MLDs witnessed another considerable change with the introduction of the *Collins COBUILD English Language Dictionary*. The dictionary was the first to fully rely on corpus evidence. Grammatical and lexical information was included in an extra column, not as part of the entry. The compilers aimed at making the dictionary easily comprehensible by the target language learners and reflecting a representative picture of modern English. In this regard, corpus evidence facilitated the inclusion of common senses and the exclusion of obsolete word forms and word senses. The frequency of use was the main criterion in organizing information in the entry. The dictionary changed the perspective on dictionary examples. Instead of relying on authoritative artificial examples that are not helpful to learners, the dictionary provided only corpus-based examples of word use. It used complete real sentences to explain the meaning and the typical use of a word (Sinclair, 1992). Sinclair (1996) explained how the dictionary makers enriched their corpus with materials designed for Teaching English as a Foreign Language (TEFL). This learner-intended corpus helped lexicographers identify the pedagogical value of various linguistic information.

Lexicographers further discussed the innovative corpus-driven aspects in the multiple editions of COBUILD. For instance, Heuberger (2018) pointed to presenting the frequency of words as an initiative launched in the second edition of *COBUILD.* This initiative helped learners identify the

words they should remember because they are likely to encounter them. Unlike most dictionaries, the online version of *COBUILD* presents frequency information for a large number of words. Moreover, it offers information about the frequency of using the word over the years. Distributional and frequency information is one of the major advantages of corpus methods. Moreover, the fifth edition of the dictionary granted corpus access to learners who could search a five-million-word corpus and recall the concordance of the most and least frequently used words. Concordance lines may serve as templates for the correct syntactic use of words. This feature has significant implications for the encoding performance of learners.

After successfully using corpus tools in the *COBUILD* dictionary, the use of corpora and corpus tools became established in compiling MLDs. Corpus-based or corpus-driven information has become the first step to compiling "good dictionaries". Corpora are now an essential part of the lexicographer's toolkit (Atkins and Rundell, 2008). The increasing reliance on corpus analysis in lexicography was parallel to the rapid developments in corpus linguistics, which witnessed several developments since the compilation of *COBUILD* in the 1980s. Dictionary makers started to develop their own machine-readable resources to elicit lexicographic information, especially for learners' dictionaries (Cowie, 2007).

Nowadays, simple definitions, grammatical information, usage notes, corpus-based examples, frequency of use and pronunciation are core features in MLDs. Table 1 lists the most notable English MLDs along with their innovative features.

Table 1. The most notable English MLDs, their date of publication and most prominent features

Year	Dictionary	Prominent features	Current practice in MLDs
1935	*New Method English Dictionary*	Controlled defining vocabulary	Maintained (except COBUILD)
1938	*A Grammar of English Words*	Inflections, derivations, collocations, phrases	Maintained
1942	*Idiomatic and Syntactic Dictionary*	Syntactic information Simple paraphrases	Maintained
1948	*A Learner's Dictionary of Current English*	Oxford's reproduction of the *Idiomatic and Syntactic Dictionary*	
1952	*The Advanced Learner's Dictionary of Current English*		
1954	*A Guide to Patterns and Usage in English*	Verb and adjective complementation	Partly maintained
1974	*Oxford Advanced Learner's Dictionary*	Oxford's reproduction of the *Idiomatic and Syntactic Dictionary*	
1978	*Longman Dictionary of Contemporary English*	Controlled defining vocabulary Systematic grammatical codes Usage notes	Maintained
1987	*Collins COBUILD English Language Dictionary*	Corpus-based Authentic examples Complete sentence definitions Grammatical and lexical information in a separate column Frequency-based sense order	Maintained in COBUILD
2002	*Macmillan English Dictionary for Advanced Learners*	Corpus-based Controlled defining vocabulary Grammatical information	Maintained
2008	*Merriam-Webster's Advanced Learner's English Dictionary*	American vocabulary and usage	Maintained in Merriam-Webster

2.1.2 New Types of Dictionary in the Electronic Era

The field of lexicography has witnessed major changes over the years. In his description of the significant influences in the field over two millennia, Hanks (2013) considered the invention of printing the first influence, with computers and corpora the second factor. The new computational and programming technology gave lexicographers and users innovative tools to compile and

consult dictionaries. Further, it altered the traditional category of dictionary users by including machines in addition to humans. As a result, lexicographers have new demands to meet in their dictionaries.

Moreover, De Schryver referred to the accessibility of the internet, the availability of effective search engines and massive databases as game-changers in lexicography. The online resources available to language users and learners contain vastly more information than that offered by any dictionary. Therefore, he chose *Google Search Engine* as the best lexicographic resource (De Schryver, Chishman and da Silva, 2019). This untraditional choice reflects how the perspective on the typology of dictionaries and what is classified as a lexicographic resource is developing over time because of technological developments.

Jackson (2018) summarized categorizing criteria that apply specifically to online dictionaries. Before the electronic era, dictionaries were institutional (i.e., the result of a systematic work published by academic publishers). Then, a new type of dictionary was introduced based on collective efforts that were not compatible with academic institutional work (e.g., *Wiktionary* and the *Urban Dictionary*). Recently, aggregate dictionaries represent a new electronically motivated type of dictionary. They offer access to multiple dictionaries when a word is searched.

Verlinde and Peeters (2012) explained how dictionaries have changed in the electronic era. Users are able to consult various dictionaries in the same search process through meta-dictionaries. Given the availability of the internet to most users, meta-dictionaries provide quick and easy access to the contents of multiple dictionaries simultaneously. However, the challenge of deciding the appropriate amount of information is more pressing in such cases. Meta-dictionaries are either aggregators or portals. Whereas portals are websites that host hyperlinks to several dictionaries, aggregators paste the contents of other dictionaries to the same website (Dziemianko, 2018). The following are some of the results of looking up *inference* in an aggregator (*The Fine Dictionary*) and a portal (*OneLook*).

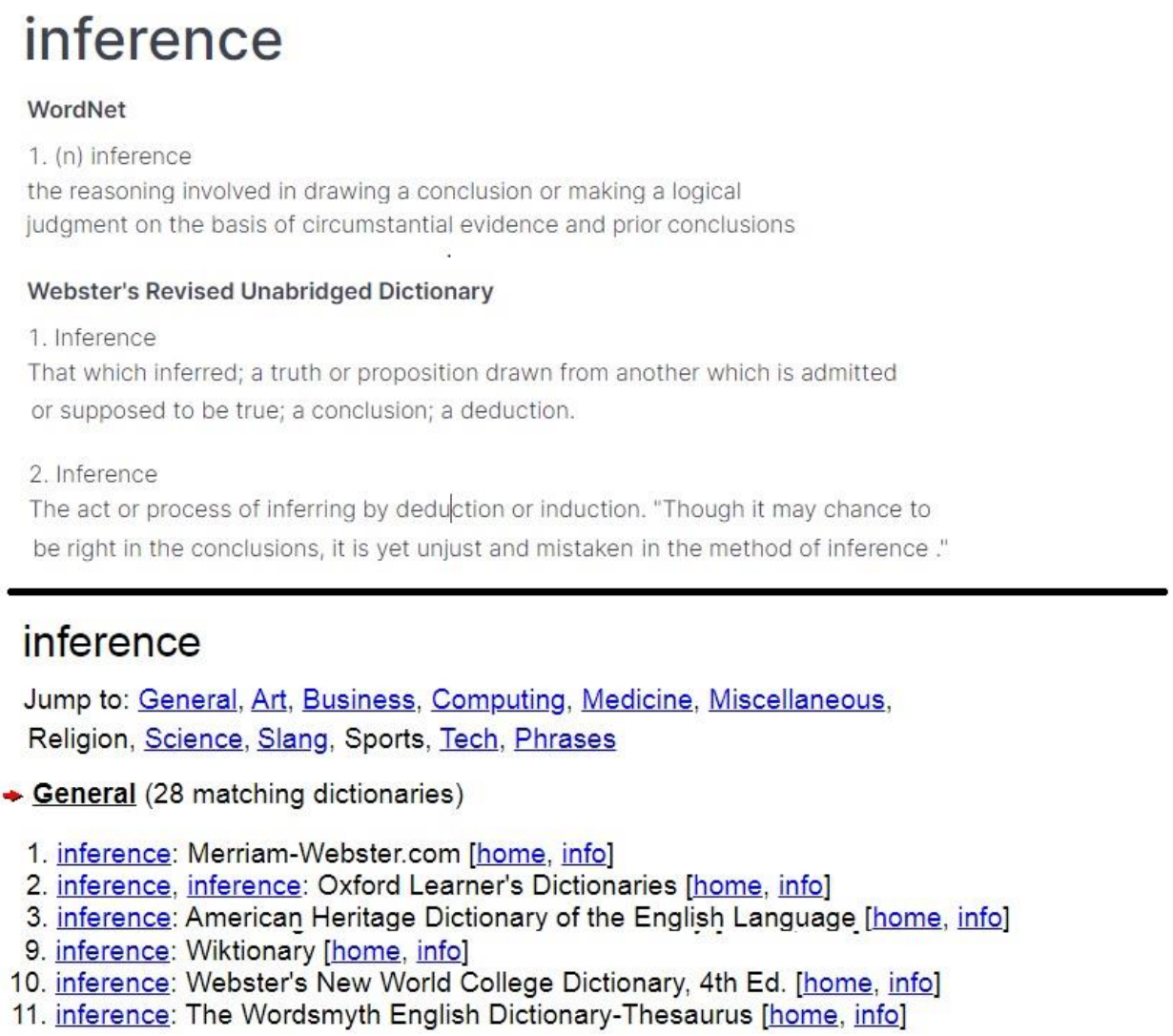

Figure 1. *Inference* in *The Fine Dictionary*[7] (aggregator) and *One Look*[8] (portal)

The Fine Dictionary retrieves definitions from a lexical database (WN) and encyclopedic dictionaries. It also cites information about the etymology of the word from the same lexicographic resources. The dictionary also collects quotes in which the target word is used by famous authors. Quotations are different from the usage of the word, which offers genre-classified citations for the target word. *The Fine Dictionary* presents some of the lexical information (e.g., synonyms and hypernyms) in WN under the label "related words". The

[7] The dictionary is available at https://www.finedictionary.com/ (accessed on 1-12-2022).

[8] The dictionary is available at https://www.onelook.com/ (accessed on 1-12-2022).

dictionary presents authentic information from four lexicographic resources, but it re-classifies the information.

One Look retrieves entries for *inference* from 51 dictionaries. It also provides hyperlinks to similar words, collocations and usage examples. In addition to the rich syntagmatic and paradigmatic information, *OneLook* hosts links to the idioms which are relevant to the meaning of the target word (not necessarily containing the target word), words that rhyme with the target word and coined words relevant to the target word. *OneLook* follows mainly a topic-based categorization of the results (e.g., art, business, computing), although sociolinguistic (e.g., slang) and linguistic (e.g., phrases) classifications are also present. Such features are not, at least collectively, present in other dictionaries.

Despite the innovative features of *OneLook*, several drawbacks impede beginners and intermediate users from making the best use of this portal. In addition to the challenge of overloading the user with information, *OneLook* displays the results of different lexicographic resources without differentiation between dictionaries (e.g., OALD), encyclopedias (*Wikipedia*) and databases (e.g., WN) or institutionalized and non-institutionalized dictionaries. MLDs such as OALD and *Collins COBUILD* are displayed under the label "General", as well as collaborative dictionaries such as *Wiktionary* and other lexicographic resources (e.g., *Wikipedia*). This clustering of heterogeneous lexicographic resources may perplex beginners and intermediate users who may consider the information in non-institutionalized dictionaries as reliable as the information in institutionalized ones. The same is applicable to the other features of this portal. Whereas related idioms include authentic linguistic data such as *read between the lines,* "related invented words" is a new tool designed to generate words based on semantic similarity and phonotactics. The invented words presented by *OneLook* do not reflect the linguistic reality of the target word.

Thanks to the availability of the internet, language users have become more involved in the lexicographic process as contributors. Meyer and Abel (2018) explained how the active role of dictionary users changed from giving feedback to editing entries and ended with crowdsourcing dictionaries. They referred to the early attempts to involve users in dictionary compilation in the 19th Century. The compilers of OED invited English readers to send evidence of the use of target words in the literature. Expert lexicographers revised the user-generated content during the synthesis of the dictionary entries. They also discussed the taxonomy of collaborative dictionaries

based on the role of users, publishers and editors. First, collaborative-institutional dictionaries refer to the crowdsourced entries which are generated by users and checked and edited by institutionalized publishers. They are neither the output of professional lexicographers nor chaotic entries written by reckless users. *Macmillan Open Dictionary* is an example of this type. Second, semi-collaborative-institutional dictionaries are based on the contribution of specialists in a certain field who submit their work to editors. Professional editors review the expert-generated content without making extensive modifications. Third, open-collaborative dictionaries represent a new paradigm in lexicography. Such dictionaries are created by users and are not subject to editorial work. Despite the questionable reliability of such dictionaries, they are successful in capturing neologisms, detecting new word uses and reflecting a variety of sociolinguistic information (Meyer and Gurevych, 2012).

Institutionalized online dictionaries acknowledge the active role of users. *Collins COBUILD*, for instance, provides a "suggest this word" form, and *Macmillan* offers a similar "add a word" form. The supervised use of crowdsourcing helps lexicographers cope with the dynamicity of the language without vandalizing the dictionary's content. *Macmillan* encourages users to check whether the word is already in the dictionary, avoid words users invented and search for evidence for word usage before submission (although the last condition is optional in the submission form). Table 2 lists some of the submitted words to *Collins COBUILD*, their submission date, status and inclusion in other dictionaries. The status of the words and their inclusion in, or absence from, other dictionaries show the lexicographic controversy over what can be accepted as a headword. The phrase *quiet quitting*, for instance, is listed as a headword in *Collins COBUILD,* but other MLDs have not admitted it yet. *Nitwit* is rejected by *Collins COBUILD* lexicographers but admitted in collaborative-open and collaborative-institutional dictionaries (i.e., the *Urban Dictionary* and *Macmillan Open Source*).

Table 2. Sample of the user-suggested content in the online version of *Collins COBUILD*[9]

Headword	Submission date	Status	Inclusion in other dictionaries
Covidiot	2020	Pending investigation	√ UD, MOD
Megachurch	2022	Rejected	√ UD
Nitwit	2022	Rejected	√ UD, MOD
Nontroversy	2022	Pending investigation	√ UD
Quiet quitting	2022	Published	√ UD
Warm bank	2022	Published	x

UD: *Urban Dictionary*[10] MOD: *Macmillan Open Dictionary*[11]

As mentioned above, *Macmillan Open Dictionary* is a successful implementation of crowdsourcing in collaborative-institutional dictionaries. It is evident that the institutionalized part of the dictionary presents consistent definitions of the target words and displays the same type of information for every headword. On the contrary, the collaborative-institutional version provides headwords, definitions (which vary from a single word to complete sentences), and occasionally example sentences. Such inconsistencies in the amount and type of information about headwords in collaborative-institutionalized dictionaries are even more salient in open-collaborative dictionaries such as the *Urban Dictionary* and *Wiktionary*. Figure 2 compares entries from the online *Macmillan Dictionary* (institutionalized), *Macmillan Open Dictionary* (collaborative-institutional) and the *Urban Dictionary* (collaborative-open). The lack of editorial effort is salient in the Urban Dictionary, manifested in repeated entries for the same headword, entries for non-words and typos, and unsystematic definitions with grammatical and semantic errors. However, it is also a rich repertoire of sociolinguistic and linguistic data that is not available in institutionalized dictionaries.

[9] The dictionary is available at https://www.collinsdictionary.com/ accessed on 1-15-2023.

[10] The dictionary is available at https://www.urbandictionary.com/ accessed on 1-15-2023.

[11] The dictionary is available at https://www.macmillandictionary.com/open-dictionary/ accessed on 1-15-2023.

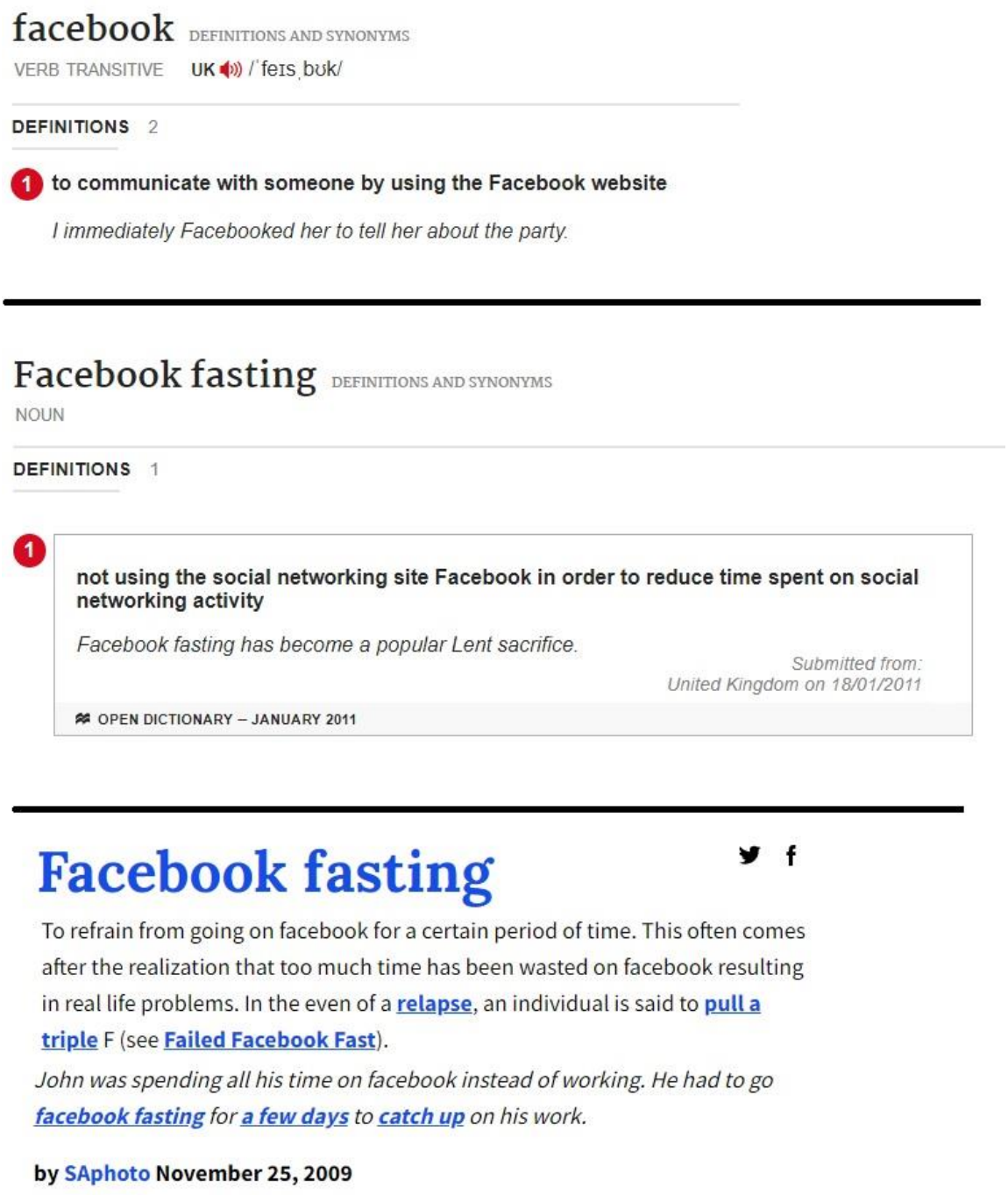

Figure 2. *Facebook* in *Macmillan Dictionary*[12], *Facebook fasting* in *Macmillan Open Dictionary*[13] and *Urban Dictionary*[14]

[12] The dictionary is available at https://www.macmillandictionary.com/ accessed on 1-15-2023.

[13] The dictionary is available at https://www.macmillandictionary.com/open-dictionary/ accessed on 1-15-2023.

[14] The dictionary is available at https://www.urbandictionary.com/ accessed on 1-15-2023.

2.1.3 New Types of Lexicographic Resources

The new lexicographic needs in the electronic era expanded to Natural Language Processing (NLP). Therefore, the scope of dictionary users enlarged to include machines, and dictionaries became important to machine learning. As this new use emerged, it was facilitated by the transition to the digital era, which led to the conversion of print dictionaries into CD format and online internet-hosted versions. Machine-readable versions of traditional lexicographic resources, such as the *Longman Dictionary of Contemporary English* (LDOCE), have been effectively employed in automatic word sense disambiguation (Agirre and Edmonds, 2007). In addition, some lexicographic resources, such as FrameNet, have been constructed with human users and machines in mind (Baker, Fillmore and Lowe, 1998).

Baker, Fillmore, and Lowe (1998) constructed a new type of lexicographic project entirely dependent on the Frame Semantics theory and computer-aided analysis. The FrameNet (FN) project was designed for two user categories, humans and machines. FN was motivated by Fillmore's and Atkins' (1992) perspective on imitating the mental lexicon in a frame-based dictionary. FN is aimed at providing users with all the information available to a native speaker when they use a word. Therefore, FN essentially states information about the typical situation in which a word sense is used, the typical participants in such a situation and the valence description of a word sense. It also lists several annotated examples to explain how each word sense is utilized in naturally occurring contexts.

Given the novelty of FN, the eight properties introduced by Atkins and Rundell (2008) to identify the type of dictionary are not straightforwardly applicable. FN is a monolingual resource covering general language, although several multilingual versions of FN have been introduced. As for the size, it is a dynamic, continuously updated project. Its current coverage of lexical units is more limited than pocket dictionaries (13,685 lexical units are included in the current version). FN is accessible online as a browsable web-based database. The organization of information in FN combines word-to-meaning, concept-to-word and thesaurus-like features. FN has a lexical unit index and frame index and places related words in the same frame. FN identified its target users as teachers and students of linguistics and scholars in natural language processing. Therefore, it has been widely used in multiple natural language processing tasks. Finally, FN can be used for encoding and decoding purposes because of its lexico-grammatical features.

According to recent taxonomies, FN is a poly-function lexicographic tool that serves communicative, operative and cognitive functions.

WordNet (WN) represents another new type of lexicographic project that is aimed at reflecting the mental lexicon of native speakers and utilizing technological tools. WN is a linguistic ontology with a lexicon of word senses and an ontology linking concepts through semantic relations. It organizes words into synonym sets (synsets) and provides a brief definition (gloss) for each synset. Members of the same synset represent the same concept and can replace each other in a given context without changing its truth value. WN creators used synonymy, hyponymy, meronymy and antonymy, among others, to structure the database (Prevot et al., 2010; Speranza and Mognini, 2010). First, WN uses synonyms as indicators of the intended sense of a polysemous word. Using *plank* in the same synset with *board* represents the sense denoting a piece of lumber. In contrast, the use of *committee* indicates another sense of *board* (i.e., a group of people with supervisory power). Then, glosses were added to synsets to improve the disambiguation process (Miller et al., 1993).

In WN, words are placed in part-of-speech nets, and each net has its organizing principles. The net of nouns places them according to general taxonomic features. Each noun is ultimately traced back to the top node "entity", but nouns differ across lower ontological levels. Nouns can belong to living or non-living entities. Living entities are animals, plants or persons; non-living entities are artifacts or substances, among others. Nouns are defined according to their superordinate terms and distinguishing features. The superordinate term covers several synsets while the distinguishing features (e.g., attributes, functions and parts) differentiate one synset from another (Miller, 1998).

Like FN, the traditional categorizing features of dictionaries do not directly apply to WN. WN is still a poly-function tool that can serve operative, cognitive and communicative purposes. It is also more frequently used in machine-learning contexts than in human-consultation situations. The WN approach maximizes homonymy. However, Martínez Alonso (2013) referred to the criticism directed to databases such as WN for listing word meanings regardless of their relatedness as a set of unrelated definitions.

In summary, the new advances in lexicography increased the scope, use and users of dictionaries. Moreover, they resulted in the creation of innovative lexicographic projects. Table 3 clusters potential users, uses and types of lexicographic projects in modern lexicography.

Table 3. Various users, uses, types and examples of dictionaries

Dictionary users	Dictionary uses	Dictionary types	Examples
Adult, non-native, language learners	Understanding word meaning,	Monolingual,	*Oxford Advanced Learner's Dictionary of Current English*,
	Using a word correctly,		*Longman Language Activator*,
	Knowing the equivalent word in a foreign language	Bilingual, Bilingualized	*Cambridge's English-Arabic Dictionary*
Adult, native or non-native, Field experts	Gaining knowledge about a topic	Specialized	*Merriam-Webster's Law Dictionary*
Language teachers, Applied linguists	Teaching correct word use, teaching vocabulary	Dictionaries of idioms, collocations	*Online Oxford Collocation Dictionary*
Advanced, native speakers, field experts	Understanding word meaning,	Monolingual, general language or encyclopedic dictionary, Scholarly dictionaries	*Encyclopedic Dictionary of the Bible*, *Oxford English Dictionary*
	Knowing the etymology of a word,		
	Gaining knowledge about a topic		
Linguists in the NLP field	WSD, Semantic Role Labeling, Question-answer tasks	Electronic/ machine-readable database	*FN*, *WN*

2.2 Dictionary and Databases Structures

The structures of conventional dictionaries are discussed to highlight the differences in the content and interface between FN, WN and conventional MLDs. Comparing the structures of the three lexicographic resources may also highlight the reasons for learners' limited use of FN and WN.

Although multiple dictionary structures (e.g., access structure, mediostructure, megastructure) exist, macro- and microstructures are the main focus of most lexicographic research. This fact may be attributed to the clarity of the binary taxonomy, its direct applicability to dictionary entries and its dependence on lexicographic and linguistic concepts. In contrast, the other mathematical taxonomy of dictionary structures is based on the Set Theory, which may not be as familiar to lexicographers and linguists.

Bejoint (1993) referred to Rey-Debove as the French lexicographer who differentiated between the macrostructure and the microstructure. The two terms denote the wordlist and the entry contents, respectively. Unlike the current situation, the microstructure was considered "rigid" and

consistent across dictionary entries. The macrostructure, however, can vary by adding or removing a word from the wordlist.

2.2.1 Macrostructure and Megastructure

Bogaards (2013) criticized many published reviews of dictionaries because they usually focus only on the macrostructure to appraise or criticize the new words or word uses in the wordlist. The macrostructure of a dictionary refers to all entries of the lemmas in the wordlist. In this regard, corpus-based frequency, user needs and dictionary types play a significant role in the choice of words. Although frequency lists are now essential to compile wordlists for dictionaries, user needs and dictionary types have more important roles. For instance, a synonymy dictionary would discard a frequent word if it does not have a synonym.

Atkins (2008) specified the types of wordlist as either comprehensive, an unattainable goal, or selective, the standard decision. She further mentioned the possible options for a headword list (e.g., totally homographic headword lists would grant homonyms separate entries). Also, decisions relevant to selecting lexical items that will be granted headword status are relevant to the macrostructure of a dictionary. Including phrasal verbs, multiple-word expressions, and morphological forms in the headword list differs across dictionaries. In addition, the organization of words may be alphabetical or thematic. The macrostructure of a dictionary also involves the overall type of dictionary entries (i.e., flat or tiered). Wiegand, Beer and Gouws (2013) considered the macrostructure "responsible for the order of all elements of a printed dictionary". This definition intersects with Bogaards' (2013) definition of the "megastructure". The generic label "megastructure" refers to the connection and organization of all dictionary components (i.e., front matters, entries in the wordlist and back matters). Adamska-Sałaciak (2013) stated that the megastructure of a dictionary combines its macrostructure and outside matters.

At the macrostructure level, any MLD defines a considerably larger number of headwords than FN and WN. WN includes 117,000 synsets, whereas FN contains 13,686 lexical units. In addition, adding new words in FN, for instance, aims to enhance the database's coverage with words that have already been in use. It does not reflect the dynamicity of the language by any means. On the contrary, newly-added headwords to conventional dictionaries are usually neologisms that correspond to real-life changes (e.g., *Brexit*, *vax*, *long-Covid* in OALD).

2.2.2 Microstructure and Mediostructure

Atkins (2008) explained that the microstructure of a dictionary concerns the selection and presentation of information in the entry. Lexicographers may include or discard specific pieces of etymological, external and internal information about a word. Internal word information is relevant to its morphology, orthography, semantics and phonological features. Internal information embraces facts relevant to both the word's form and sense. External information refers to the relation between a word and other words. It includes paradigmatic relations (e.g., POS and synonyms), syntagmatic relations (e.g., collocations), relational links (e.g., cross-references to derivational forms) and usage information (e.g., genre, dialect).

Unlike the linguistic-based taxonomy, the use of cross-reference to any relational, paradigmatic or illustrative information is called "mediostructure" by scholars following the Set-Theory-based typology. The term refers to all information relevant to the entry but placed elsewhere and cross-referenced in the entry (Bogaards, 2013).

Figure 3 displays the lexical entry of the same sense of *smart* in the online version of *Collins COBUILD*, FN and WN. As displayed, the microstructure of the three projects differs in data description and presentation. *Collins COBUILD* displays information about the target word's frequency, pronunciation and morphological forms. It also provides sociolinguistic details on word use, e.g., "mainly British". It structures the data in a traditional dictionary format and uses the online features available to dictionaries such as hyperlinks, audio and visual options. The FN entry mentions only two pieces of conventional dictionary information: the definition and POS. As an innovation, it identifies the frame evoked by the LU and describes which FEs occur with the LU and how they are syntactically realized. In addition, it tabulates the three-layered valence patterns of the target word and includes annotated examples. WN focuses on the synonyms and the antonyms of an adjective and lists other adjectives that are semantically similar to them. Usefully, WN records the morphological forms that are derivationally related to the adjective.

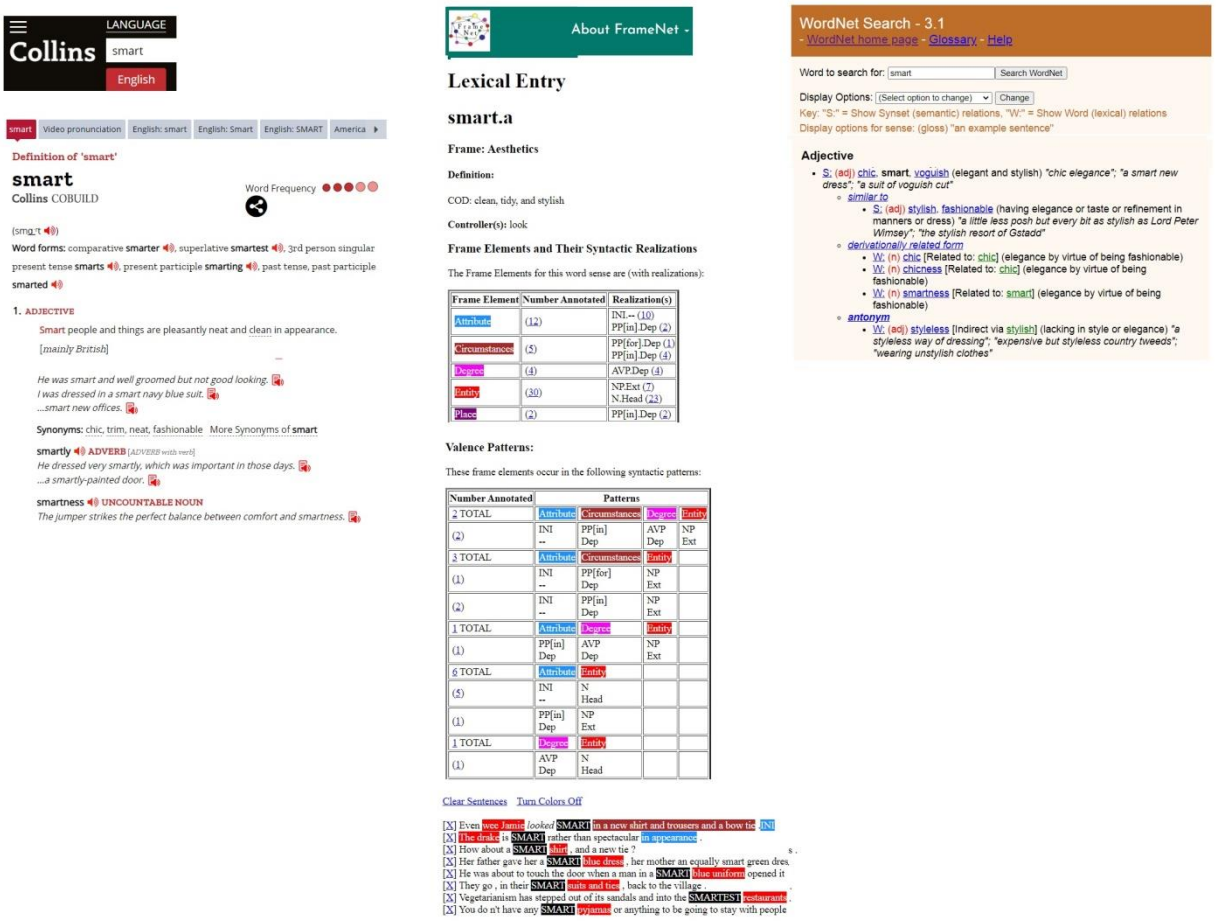

Figure 3. *Smart* in *Collins COBUILD*[15], FN[16] and WN[17]

2.2.3 Accessibility

Gouws (2018) considered the accessibility of information in a dictionary the main factor of its success or failure. Accessibility should be the primary concern of theoretical and practical lexicographers alike. Rapid access and information retrieval should be the main criteria according to which lexicographers design the macrostructure of a dictionary. A user would successfully consult a dictionary based on information accessibility, retrieval and clarity. The traditional alphabetical ordering of words in the macrostructure provides users with a single alphabetical

[15] The dictionary is available at https://www.collinsdictionary.com/dictionary/english accessed on 1-2-2022.

[16] The dictionary is available at https://framenet.icsi.berkeley.edu/fndrupal/ accessed on 1-2-2022.

[17] The dictionary is available at http://wordnetweb.princeton.edu/perl/webwn accessed on 1-2-2022.

access structure that starts from the top of a page and moves downwards. A nested access structure increases the density of information in the macrostructure and, accordingly, slows the accessibility of information. Technology granted online dictionaries better access structure than print dictionaries. Online dictionary users do not have to follow a downward route to find the item they want. Users instead type a search string, and all relevant information should be automatically retrieved. This grants easier and quicker access to information. In addition, it allows rapid access to information at the microstructure level if the online dictionary is well-planned and makes full use of the available technology.

Wiegand, Beer and Gouws (2013) differentiated the outer access structure from the inner access structure. Whereas the former gives access to external data in the macrostructure of the dictionary, the latter guides the user towards a particular piece of information in the accessible entry. Dziemianko (2018) regarded the simplified inner and outer access structures in electronic dictionaries as remarkable improvements in lexicography. They speed up the lookup process and improve its accuracy. Enhancing the outer access of dictionaries is reflected in the developed findability of a headword, inflected forms and MWEs. Looking up a homonymous word in an electronic dictionary provides users with "step-wise outer access" to the different words realized by the same orthographic form. It allows users to access the desired information through hyperlinks. Similarly, the inner access is simplified by using hyperlinks to the additional information a user may be interested in and controls the amount of information displayed simultaneously. This feature is especially useful in polysemous entries.

Despite the consensus over the improved accessibility of dictionaries, disagreements exist over the relationship between the positive and negative influence of the macrostructure on information accessibility. Although Gouws (2018) recommended an alphabetical access structure for users with relatively less advanced dictionary experience, Chi (2013) listed many factors that impede the accessibility of information in an alphabetically structured dictionary. Chi (2013) explained the influence of different macrostructural choices on the accessibility of information by ESL learners. The study focused on the *Oxford Advanced Learners' Dictionary* (OALD), which used to follow the non-alphabetical ordering of the wordlist in the first six editions. OALD followed the nested ordering of words according to morphological and semantic relations. It is argued that the non-alphabetical order helps learners access information needed for encoding purposes, enriches their vocabulary and allows them to realize semantic and morphological connections in

the lexicon. After the sixth edition, the dictionary compilers adopted the alphabetical organizational principle in the macrostructure to increase the speed and ease of information accessibility. This change reflected an increasing interest in decoding purposes over encoding ones. However, the alphabetical ordering may impose several challenges on ESL learners whose mother tongue is not Latin-based and decrease the findability of a translation word. Given the diversity of languages and cultures of learners, an alphabetically ordered learners' dictionary cannot provide a customized macrostructure that suits all learners. Therefore, the study suggested a thematic ordering or a synonymy-set ordering of MLDs to enhance the accessibility of information in the macrostructure. At the microstructure level, signposts and menus improve the accessibility of information in the entry (e.g., finding the correct sense in long entries). Heuberger (2018) also reiterated the influence of signposts on improving the access structure of MLDs. Signposts can be synonyms, hypernyms or short definitions that help users locate the sense they want in long dictionary entries.

Humbley (2018) discussed another factor that might determine the choice of alphabetical or thematic order of words in a dictionary. Whereas alphabetical order is the norm in general language dictionaries, thematic order is preferred in specialized dictionaries that deal with terms in a specific field. Notwithstanding, an additional alphabetical index of words is also needed in thematically-organized dictionaries to improve the accessibility of information to less experienced users. The flexibility of presenting information in electronic dictionaries allows lexicographers to provide users with hyperlinks to access the thematic structure of the dictionary.

The online version of OALD enhances the accessibility of words through different features besides the conventional search box. First, it offers direct access to various word lists based on the user's needs. Oxford 5000, for instance, includes the most frequent and learner-relevant words, whereas the Oxford Phrasal Academic List (OPAL) consists of the most frequent words in academic English. Tóth (2019) states that OALD's wordlists are regularly updated. Oxford 3000, for instance, usually changes every five years, in line with the publication of the printed learner's dictionary. Oxford 3000 is the recent implementation of the calls for using controlled defining vocabulary. The concept was implemented in the fifth edition of the dictionary (OALD5) in 1995, and the core vocabulary list was named Oxford 3000 in the seventh edition (OALD7). The recent changes in Oxford 3000 omitted 907 words, including some derivational forms (e.g., actively, calmly, enjoyable), phrasal verbs (e.g., come across, come down), and serial numbers.

The changes included the addition of words relevant to information technology (e.g., *blog*, *download*, *update*), education (e.g., *assessment*, *graduate*, *illustration*), traveling (e.g., *airline*, *helicopter*), nature (e.g., *earthquake*, *hurricane*), sports and entertainment (e.g., *golf*, *laughter*, *ski*). A corpus-based investigation of a sample of the added and deleted words showed how the word frequency when compiling the list affects the word's inclusion or exclusion (Tóth, 2019). In this regard, one of OALD's wordlists provides direct access to newly added words and phrases[18]. The same categories detected by Tóth (2019) seem to be present in the list of newly-added headwords, such as *Tweeter*, *airplane mode*, *crypto* and *digital currency*. Second, OALD enables thematic access to headwords (i.e., topics). "Topics" is used as a label at the microstructure level and as a separate categorization to access words thematically at the macrostructure level.

On the contrary, synsets in WN can be accessed only through the search box. FN provides direct access to the lexical units through the alphabetical LU index and indirect access through the frame index, in addition to the search box. Therefore, the access structure of OALD is significantly simpler and more informative to users than those of FN and WN.

2.3 Word Meanings in Dictionaries and Word Uses in Corpora

This section discusses the differences between word senses in dictionaries and word uses in corpora. It overviews the challenge of associating word senses with word uses and the more serious challenge of deducing meanings from word uses.

Discovering senses from corpus citations does not follow a conventional method. Although corpus tools present numerous authentic word usages, converting corpus citations into an organized list of senses that appeal to dictionary users is a laborious lexicographic task. The final list of senses differs from one dictionary to another according to editorial policy, target users and available resources. Therefore, despite aiming at an objective representation of linguistic reality, dictionary senses do not correspond to it (Lew, 2013).

Kilgarriff (1997, 2007) explained how lexicographers might extract meanings from the patterns of use they find in a corpus. He advocated the replacement of the genus-differentiae models of sense separation with corpus-driven models. Lexicographers used to choose a genus that was general enough to include the multiple senses of a word; the differentiae were then identified to

[18] The list is available at https://www.oxfordlearnersdictionaries.com/wordlist/new_words.

separate the multiple senses from each other. Lexicographers relied on introspection and semantic analysis to identify both. According to the corpus-driven model, word senses are abstractions derived from clusters of corpus citations. Lexicographers first examine the concordance of the target word. Second, they cluster the sentences according to the similarities between them. In the most challenging step, which may not be entirely explicit even for lexicographers, they think about the criteria for grouping the members of each cluster. Clusters vary according to the corpus, the purpose of clustering and the user. Clustering citations display similar usage patterns and, arguably, similar meanings. If the uses of the word appear in totally separate clusters, the word is hypothesized to be ambiguous. However, if the clusters of uses are not well-defined, the word is characterized as vague. After clustering, the lexicographer, or the linguist, starts the qualitative analysis of the similarity patterns and differences entailed in the clusters. A pattern is elevated to the status of a dictionary sense if it has a sufficient frequency and its meaning is not easily predicted from the standard meaning and use of the word. Lexicographers have to make decisions about the distinctiveness of a pattern from other uses and its embeddedness in the shared knowledge of native speakers before recording it as a separate dictionary sense. Finally, they convert their remarks and findings into dictionary definitions.

Kilgarriff (2005) proposed another model that aimed at putting corpus into dictionaries (PCID). Dictionary entries expose differences in meaning and do not concentrate on differences in forms. Corpora saliently display different uses of word forms, but meaning differences are present only implicitly in the uses. The PCID model is an intermediate level linking dictionary senses to corpus uses for word sense disambiguation purposes. The model relies on collocate-to-sense mapping. A wide range of collocates or sentential patterns is retrieved from a corpus for each word. Collocates are associated with an individual dictionary sense. Therefore, the associated sense can be automatically assigned if the collocational pattern is encountered in a future sentence. The PCID model adds a novel grammatical dimension to the collocational relation. It uses a triplet of W_1RW_2 relation that enhances the performance of WSD. Unlike clustering, collocate-to-sense mapping would not be considerably affected if the corpus was changed or discarded. Using a target word in a collocational pattern made the clustering process more manageable and enhanced the possibility of automatic WSD (Kilgarriff, 1992). Moreover, suppose the collocational pattern has an idiomatic meaning. In that case, the ambiguity of a word

is automatically resolved because, in the absence of the idiomatic phrase, the idiomatic meaning is excluded from the context of interpretation (Britton, 1978).

Kilgarriff and Rundell (2002) outlined the use of collocational patterns displayed in word sketches in sense identification. The analysis of the word sketch of *challenge* demonstrated how the lexico-grammatical behavior of a word mirrors its meanings. Words appearing in the grammatical pattern *challenge prep[to] noun* pointed to two *challenge* types. The first is a *challenge to prevailing ideas,* while the second is a *challenge to dominant power and authority*. Collocates participating in the object of relation (i.e., *verb challenge*[*object*]) indicated several meanings, such as an item constituting a challenge and an agent initiating a challenge. Figure 4 shows some of the options offered by word sketches.

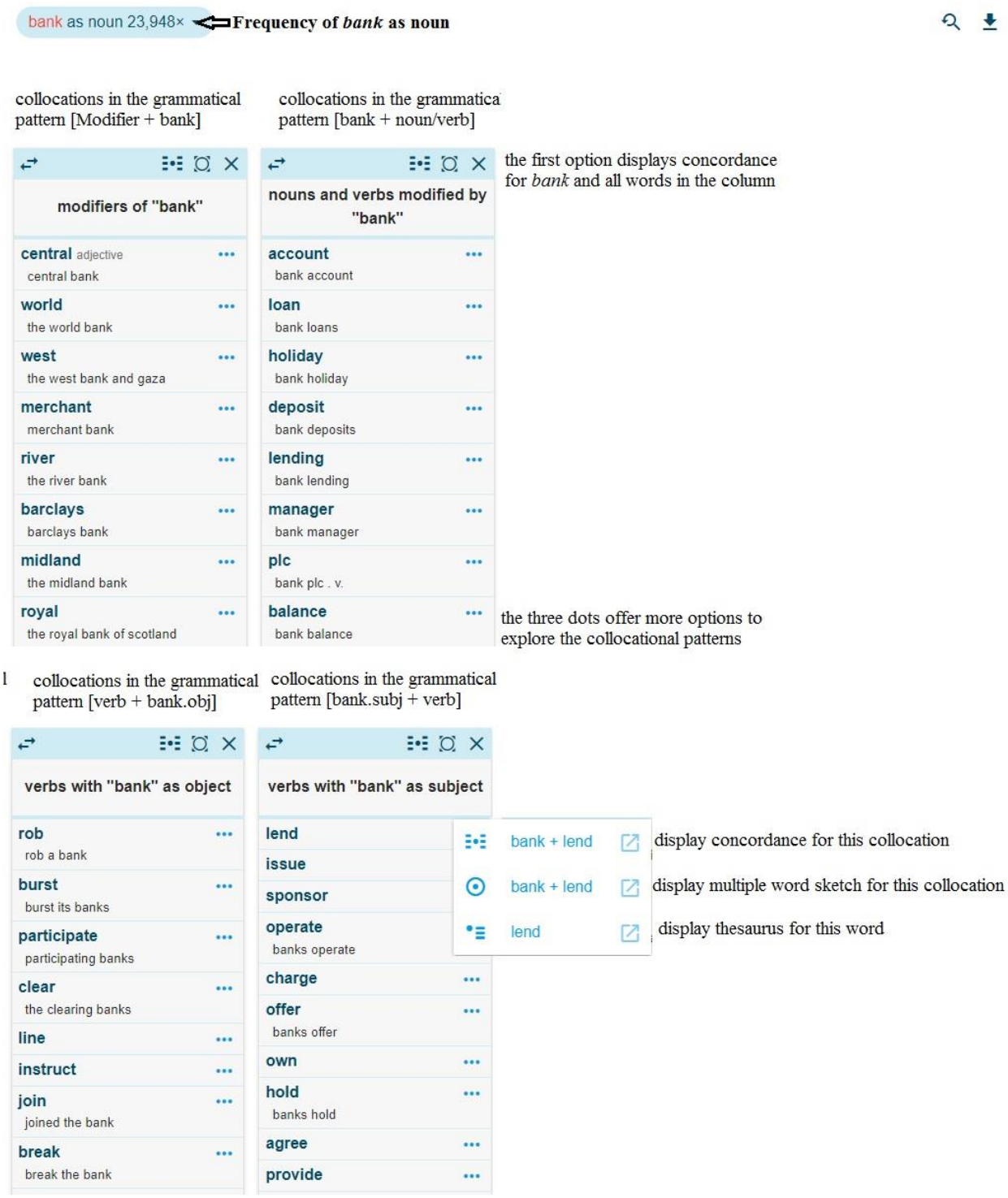

Figure 4. The word sketch for *bank* in the BNC

Smirnova (2021) used collocational patterns as sense distinguishers and polysemy detectors. However, the study did not rely on a theoretical linguistic background. It was instead motivated by the literature on psychology. The study analyzed a sample concordance of the noun *awe* cited

from a 14-billion-word corpus. The collocations of the target word and the concordance helped the scholar define the multiple senses of *awe* and the evaluative attitude of the experiences expressed by the different uses of the word.

Similarly, De Schryver and Nabirye (2018b) manually annotated a sample concordance of the verb *–v-* in Lusoga. They could map the different usage patterns to meaning potential and construct two entries for the verb accordingly. Senses were organized according to their frequency in the analyzed sample.

Hanks (2004) revealed through his corpus pattern analysis how the meanings of a word could be mapped to patterns of usage. His project required a massive lexicographic effort to process word uses, find and record usage patterns and associate each pattern with a meaning. He adopted a corpus-driven approach based on the Theory of Norms and Exploitations to examine word meanings in contexts instead of assuming the existence of meaning in isolation from the context. The result of this approach is the dynamic project *Pattern Dictionary of English Verbs* (PDEV). Table 4 contrasts the corpus-driven presentation of *need*.v in the PDEV to the traditional representation of the lexical meaning of *need*.v in the online *Oxford Dictionary*: hosted by *Dictionary.com*.

Table 4. *Need*.v in *PDEV*[19] and *Dictionary.com*[20]

PDEV	Dictionary
1 Human | Institution need Eventuality | Entity [[Human | Institution]] requires that [[Eventuality | Entity]] must be realized or available, in order to accomplish some goal	to have need of; require:
2 Entity 1 | Eventuality 1 need Entity 2 | Eventuality 2 [[Entity 1 | Eventuality 1]] is an essential precondition for or attribute of [[Entity 2 | Eventuality 2]]	to be under an obligation
3 Entity | Eventuality need to+INF An essential precondition for the realization of [[Eventuality]] is that [V] must be realized	to be in need or want.
4 Human need to+INF [[Human]] must do [V]	to be in need or want.
5 Plant | Animate need Eventuality | Stuff [[Plant | Animate]] must have [[Eventuality | Stuff]] in order to survive and flourish	

Dictionary identifies four meanings of *need* as a verb, whereas PDEV realized five different patterns based on 28,352 corpus instances. The four main senses in Lexico express essential

[19] The project is available at https://pdev.org.uk/ accessed on 1-2-2022.

[20] The dictionary is available at https://www.dictionary.com/ accessed on 15-1-2023.

requirement, necessity or obligation and the archaic sense of "be necessary". However, PDEV divides requirement, necessity and obligations based on the patterns. For instance, the obligation on a human to do an action is separated in pattern 4. It requires a human in the subject position and an infinitive form after *need*. This pattern is semantically and syntactically dissimilar to pattern 1, for example. Pattern 1 allows both humans and institutions in the subject position. It is the party that requires the realization of an entity or eventuality to reach a goal.

In the same vein, Litkowski (2014) aimed to describe the syntagmatic patterns associated with a preposition, given that dictionaries only list the different senses of a preposition. Exploiting available digital lexicographic projects, the *Pattern Dictionary of English Prepositions* used data from Oxford English Corpora, The Preposition Project and FN to describe the prototypical patterns of preposition usage. However, it relies heavily on the FN database.

Chapter Three – Ontological Approaches to Meaning

This chapter focuses on the theoretical foundations of WN, a linguistic ontology and a lexical database. It explains sense relations, used extensively in WN, and discusses how they are represented in lexicography. The structure of WN and the presentation of meanings in the database are discussed, too. Finally, lexical ontological databases constructed after WN and suggested relatively different solutions to splitting and lumping senses are overviewed.

3.1 Sense Relations in Lexical Semantics

Cruse (2000) advocated the conceptual approach to meaning, which relates linguistic forms to concepts. Concepts are defined as "organized bundles of stored knowledge". Special links such as "is kind of", "is part of" and "lives in" link concepts in a "complex multidimensional network". Linguistic forms, specifically lexical items, are directly linked to a concept and indirectly activate related concepts. For instance, the lexical item *horse* is directly linked to the concept HORSE and indirectly activates the concept ANIMAL through the link "is a". In this regard, words and concepts can be in a one-to-one relation. *Horse*, for instance, activates only the concept of HORSE directly.

On the contrary, the mapping between *bank*.v and the concept BANK is one-to-many because the lexical item gives access to at least two concepts. In addition, lexical items such as *pass away*, *kick the bucket* and *die* are mapped to a single concept DIE (many-to-one mapping). Although the three cases display differences in meaning, they are mapped to the same concept DIE. Accordingly, such differences are not conceptual; they are part of the word-specific properties. Whereas *die* is neutral, *kick the bucket* and *pass away* modulate the concept DIE. This distinction suggests that the word meaning is a combination of word-specific properties and concept-related ones.

Cruse (2000) further elaborated that lexical items in the mental lexicon are not randomized. They are rather structured through, for instance, sense relations. Later, he added that sense relations hold between word meanings. They have different types, and they link the meaning of every word to that of other words. He focused only on sense relations that are significant so that important generalizations regarding vocabulary structure could be made (Cruse, 2003). Cruse (2000) identified the criteria according to which sense relations are significant in structuring the

vocabulary of a natural language. First, a relation must be recurrent enough. The hypernym-hyponym relation holds between *apple* and *fruit*, *dog* and *animal*, *car* and *vehicle*, and *chair* and *furniture*, and it is applicable to hundreds of similar pairs. Second, sense relations must be discriminating. That is to say, they exclude a significant number of lexical pairs. The hypernym-hyponym relation excludes pairs like *dog* and a*pple*, *fruit* and *animal* and several other pairs. Third, sense relations are more significant if they can be lexicalized. For instance, *x is a kind of y* is a linguistic form that lexicalizes the hypernym-hyponym relation (i.e., *dog is a kind of animal*, *apple is a kind of fruit*).

3.1.1 Paradigmatic Sense Relations

Cruse (2000) divided sense relations into paradigmatic and syntagmatic. The current study focuses on paradigmatic sense relations, which Cruse described as the most studied by linguists. Paradigmatic relations include words that can fill the same syntactic slot in a sentence. Words that can fill the empty position in *I have a glass of* _____ belong to a single syntactic category and usually express a homogenous class of semantic choices. Cruse (2002) added that being a paradigmatic relation stipulates that lexical items in this relation must fit into the same grammatical slot and belong to the same semantic type. Paradigmatic relations either express identity and inclusion or opposition and exclusion.

Hyponymy is also the most fundamental relation, among all paradigmatic relations, in structuring vocabulary. It applies to all syntactic categories but is most salient among nouns (Cruse, 2002). Hyponymy is a relation between a general term (i.e., hypernym or superordinate) and a more specific term (i.e., hyponym or sub-ordinate). *Apple* is a hyponym of *fruit* and *fruit* is a hypernym of *apple.* At the extensional level, the class of *fruit* includes the class of *apples* and, at the intensional level, the meaning of *fruit* is included in the meaning of *apple*. That is to say, *it is an apple* entails *it is a fruit* but not the other way round. Hyponymy is a transitive relation between a hyponym and its indirect hypernym. If *Abyssinian* is a *cat* and *cat* is an *animal* then *Abyssinian* is an *animal*. Some exceptions are present for the transitivity relation. To elaborate, *seat*.n is a type of *furniture*.n and *car-seat* is a type of *seat*. However, it is very odd to consider a *car seat* a type of *furniture* (Cruse, 2000).

Incompatibility is an exclusion relation that usually applies to immediate co-hyponyms of the same hypernym. It holds between terms that denote classes with no shared members. *Dogs*.n and

cats.n represent two hyponyms of *animal*.n. They are incompatible because if *X is a cat, it is not a dog.* X cannot belong to the class of *cats* and the class of *dogs* at the same time. Still, not all hyponyms of the same hypernym are incompatibles. For instance, *queen*.n and *mother*.n share the hypernym *woman*, but an individual can be both a *queen* and *mother* (Cruse, 2000). Later, he elaborated on this inapplicability of the incompatibility relation among some hyponyms. In some cases, the hyponyms of the same hypernym form more than one set. To elaborate, *novel* and *paperback* are two hyponyms of *book* but they form two different sets. Hyponyms in the set, including *novel*, are incompatibles because *a book* cannot be *a novel* and a *textbook* at the same time. Similarly, hyponyms in the set embracing *paperback* are incompatibles (e.g., if X is a *paperback* it is not a *hardback*). However, hyponyms in the two sets do not display any incompatibility. A *book* can be both a *novel* and a *paperback* (Cruse, 2002).

Another significant paradigmatic relation is meronymy. Meronymy is a part-whole relation that holds between a *finger* and a *hand* or *nose* and *face*, for instance. The meronym is the part, while the holonym is the whole. Meronymy also expresses inclusion. A *hand* includes a *finger* as an individual entity, not as a class, and the meaning of *finger* incorporates the meaning of *hand* in some way. Meronymy is lexicalized as follows: *An X is part of a Y* and *a Y has an X*. If X is a part of Y and A is part of X, then it is entailed that A is part of Y.

Meronymy can reflect necessity or optionality. For instance, *fingers* are necessary meronyms of a well-formed *hand*. However, *beard* is an optional meronym of *face*. This is reflected in the well-formedness of the holonym without a meronym. Whereas *face* is perfectly well-formed without a *beard*, it is not without a *lip* or an *eye*. Meronyms also can be fully or less integral to the whole. Integrity can be detected in the normality or oddness of linking the meronym to the whole through the phrase *attached to*. If the part is typically integrated, it would be awkward to describe it as attached to, e.g., **the finger is attached to the hand* as opposed to *the handle is attached to the door* (Cruse, 2000; 2003).

Co-meronyms are distinguishable from each other through the dissimilar functions they serve and through formal discontinuity. For instance, *car wheels* and the *steering wheel* function as movement facilitators and movement directors, respectively. Discontinuity is apparent in the detachability of the wheels from the car and their relative movement. The discontinuity of a meronym from its whole can also be visual, e.g., the discontinuity between the cuff and the sleeve (Cruse, 2003). Incompatibility is also applicable to co-meronyms. In strictly logical

meronymy, if X and Z are co-meronyms of Y, then an entity A cannot be a meronym of X and Z at the same time. To exemplify, *eye* and *mouth* are necessary and integral parts of *face*. *Eyelashes* and *lips* are meronyms of *eye* and *mouth*, respectively. Accordingly, neither *eyelashes* nor *lips* can be parts of *mouth* and *eye* at the same time (Cruse, 2000).

Cruse (2017) added that the same word can be a hyponym and a hypernym. *Dog* is a hyponym of *animal* and hypernym of *spaniel.* The same is applicable to the meronymy relation. Whereas *hand* is the holonym of *finger*, it is the *meronym* of *arm.*

Synonymy is a paradigmatic relation between words that share more semantic similarities than differences. There are three degrees of synonymy which reflect how close the two synonyms are. The highest degree of similarity is present among "absolute synonyms" and is detected through a contextual approach. Absolute synonyms should manifest identical meanings and can be used interchangeably in grammatical sentences without affecting the "contextual normality". Such conditions are almost impossible to attain in natural language because there will always be some fine-grained distinctions between words in use. The second degree of similarity is between "propositional synonyms". They can be used interchangeably in a sentence without changing its truth value. To elaborate, there is a mutual entailment between *John bought a violin* and *John bought a fiddle* because they have the same truth value. However, they do not display the same degree of contextual normality. Propositional synonyms share their propositional meaning, but they differ in expressive meaning and stylistic or discourse use. The third degree of synonymy is between near-synonyms which are usually recorded in dictionaries. Near synonyms typically share their core semantic content and differ in some non-core features. This somehow blurs the boundaries between propositional and near-synonyms (Cruse, 2000). Later, Cruse (2003) proposed that the difference between propositional and near-synonyms is detectable through the *X or rather Y* test. The test suggests that some propositional information is present in Y but not in X, and the difference is not core. Near synonyms share core semantic meaning and differ in minor aspects e.g., *he was not killed he was executed*.

Opposites represent the fourth paradigmatic relation. They are also incompatibles, i.e., *if X is long, it is not short*, but they have multiple types. On the one hand, complementaries are inherently binary pairs such as *dead-alive*, *true-false* and *inside-outside*. No entity can be *dead* and *alive* at the same time. On the other hand, antonyms are gradable adjectives that refer to certain degrees on a scale of a property such as *length* or *speed*. They can usually be preceded by

a set of degree modifiers, including *very*, *so*, or *extremely*, among others. One of the antonyms expresses a value that is below the standard on the scale of the property (e.g., *slow*, *short*), while the other refers to a value that is above the standard point on the scale (e.g., *fast*, *long*). Still, some adjectives appear in the neutral middle between the two antonyms, such as *neither fast nor slow*. What distinguishes converses from other opposites, such as antonyms and complementaries, is their usability in expressing the same state of affairs. "*A is above B*" describes the same state of affairs expressed by *B is below A* (Cruse, 2000; 2003).

Despite the importance of sense relations in explaining word meaning, no traditional lexicographic resource has applied them systematically to word definitions. Syntagmatic relations such as collocations are systematically applied in collocational dictionaries. However, synonymy, rather than paradigmatic relations, is the most used relation in traditional thesauri. Otherwise, both syntagmatic and paradigmatic relations are sporadically and selectively applied in general language dictionaries. Figure 5 shows a couple of entries from two specialized dictionaries that depended on synonymy and antonymy relations. On the right side, there is a sample of the sporadic implementation of sense relations in the online version of OALD. On the left side, there is a sample of the systematic application of synonymy and antonymy relations in two specialized dictionaries.

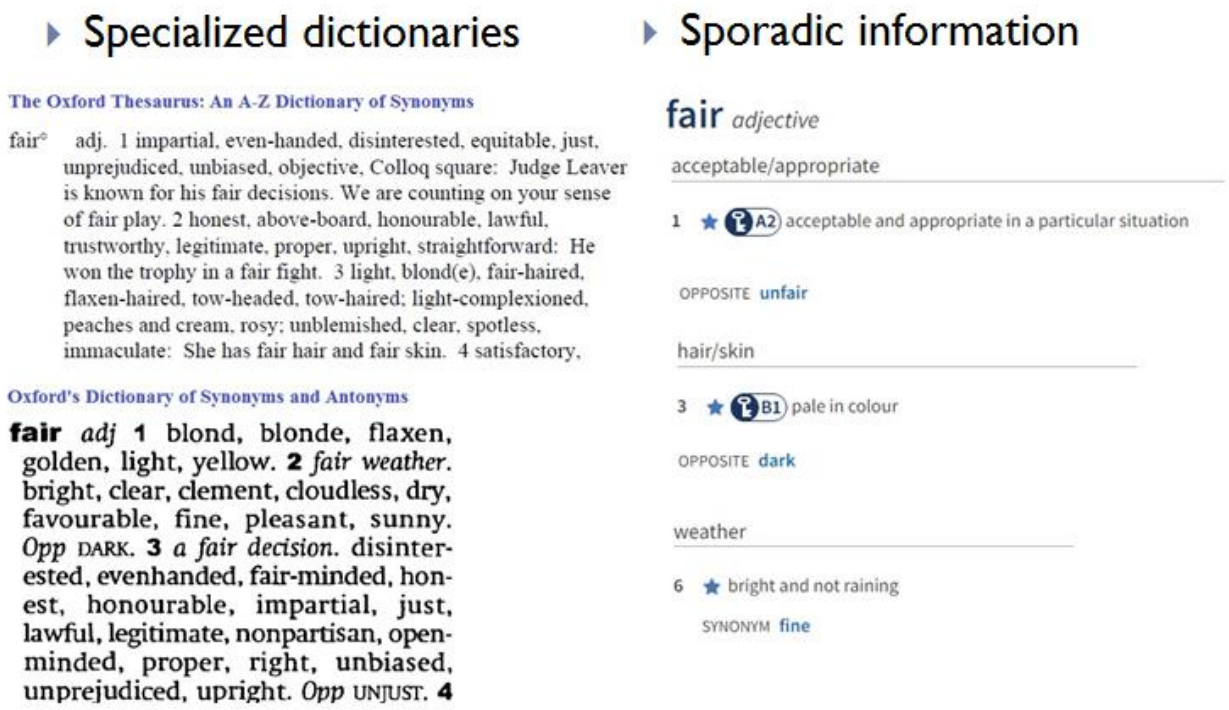

Figure 5. The implementation of sense relations in *The Oxford Thesuarus*[21], *Oxford's Dictionary of Synonyms and Antonyms*[22] and *Oxford Learners Dictionary*[23]

3.1.2 The Use of Sense Relations in Ontologies

Ontological approaches to meaning are theoretically and methodologically different from the lexicographic approaches. Lexicographers attempt to describe the linguistic reality as faithfully as possible, whereas ontologies focus on the concepts underlying linguistic expressions. That is to say, they follow the ideational theory of meaning. If several words refer to the same concept, a lexicographer can include them all as synonyms. In contrast, a formal ontology eliminates "redundant" words so that each concept corresponds to a single term. In addition, lexicographers are not allowed to judge the absence of a word from a language as a gap and are not, accordingly, allowed to invent words. Formal ontology licenses the invention of words as a strategy for including concepts that are ontologically important but are not lexicalized in some languages (Prevot et al., 2010).

[21] The entry is cited from The Oxford Thesaurus: An A-Z Dictionary of Synonyms (1st edition, p. 535).

[22] The entry is cited from Oxford Dictionary of Synonyms and Antonyms (1st edition, p.146).

[23] The dictionary is available at https://www.oxfordlearnersdictionaries.com/ accessed on 1-2-2022.

Ontologies are “specifications of shared conceptualizations”. In this regard, conceptualization is a “mental product” or knowledge extracted, by an agent, from “experience, observation or introspection”. Whereas conceptualization expresses a view of the world as conceived of by an agent, ontologies reveal which conceptualizations are shared by agents. Specification, in the ontological senses, refers to the encoding of this shared knowledge in formal or natural language representation (Prevot et al., 2010)

Linguistic ontologies are similar to dictionaries in their strong dependence on language. They resemble formal ontologies in listing concepts shared by different words in different languages. However, they are larger than formal ontologies, which usually focus on the upper levels of ontology and are not rich in linguistic expressions (Speranza and Mognini, 2010). Linguistic ontologies use language resources, including both semi-structured texts and text corpora, to derive information and extract ontological knowledge. Creators of linguistic ontologies process information in dictionaries and thesauri and propose a new organization of word senses in order to determine the concepts of the ontology. Word senses in dictionaries represent knowledge shared by a community but are presented for human users. Therefore, they usually undergo further processes of lumping (or splitting) based on the ontological purpose or the relevant natural language processing task. Ontology creators can also extract knowledge from general and specialized corpora through the use of taxonomic patterns, e.g., “x such as y”. (Lenci, 2010; Prevot et al., 2010).

The following sections overview some ontological methods for meaning representation. Although WN does not follow a precise method for sense separation, several linguistic ontologies rely on WN senses in their own processes of splitting and lumping senses. Therefore, the sporadically mentioned conventions of WN’s sense separation will be reviewed before discussing two clearer, but WN-dependent, models for sense separation.

3.1.2.1 WN Database

WN is a linguistic ontology that has a lexicon of word senses and an ontology linking concepts through semantic relations. It organizes words into synonymy sets (synsets) and provides a brief definition (gloss) for each synset. Members of the same synset represent the same concept and can substitute each other in a given context without changing its truth value. WN creators used

synonymy, hyponymy, meronymy and antonymy, among others, to structure the database (Prevot et al., 2010; Speranza and Mognini, 2010).

WN assumes the existence of a word matrix that has word meanings on the vertical dimension and word forms on the horizontal dimension. The elements or entries of the word matrix represent the association of a particular word form with a specific meaning. Synonymy is reflected in the multiple entries in the same row, and polysemy is displayed in the multiple entries in the same column. The hypothesis is visualized with examples in Figure 6. First, WN used synonyms as indicators of the intended sense of a polysemous word. The use of *plank* in the same synset with *board* represents the sense denoting a piece of lumber, whereas the use of *committee* indicates another sense of *board* (i.e., a group of people with supervisory power). Then, glosses were added to synsets to improve the disambiguation process (Miller et al., 1993).

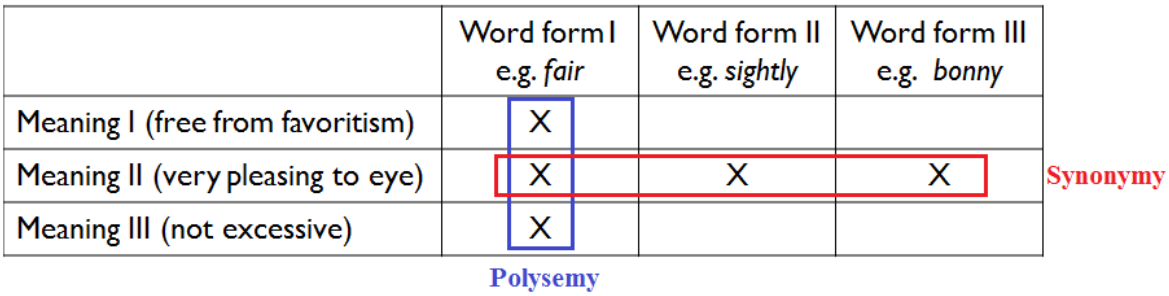

Figure 6. WN's hypothetical word matrix

WN groups words in part-of-speech nets. Each net has its organizing principles. The net of nouns places them according to general taxonomic features. Each noun is ultimately traced back to the top node "entity", but they differ across lower ontological levels. Nouns can belong to living or non-living entities. Living entities can be animals, plants, or persons, and non-living entities are artefacts or substances, among others. Nouns are defined according to their superordinate terms and distinguishing features. The superordinate term covers several synsets while the distinguishing features (e.g., attributes, functions and parts) differentiate one synset from another (Miller, 1993).

Variations in superordinate terms are explicit in WN, but the distinguishing features may be implicit or explicit. Lexicographers can anchor on both to split senses of common nouns. If a noun has different direct hypernyms, its senses are split into distinct WN synsets. *Tree*, for instance, denotes a "woody **plant**" and a "**figure** that branches from a single root". The direct

hypernyms and higher ontological levels are dissimilar in the two cases. The first sense inherits the higher structure of "physical entity", and the second sense fits within the "abstract entity" hierarchy.

Hypernym-hyponym relation is not applicable to the net of adjectives. Selectional preferences of polysemous adjectives, as well as the different antonyms of the same adjective, guided WN creators through the sense separation of adjectives. Adjective-noun patterns contribute to distinguishing nouns and disambiguating polysemous adjectives. Polysemous adjectives usually select dissimilar nouns to modify. *Short* is a value of the attribute "height" and can modify nouns referring to humans. In this case, *tall* is its antonym. Also, *short* selects nouns like *story*, *flight*, or *holiday* which do not have "height" as an attribute. However, they have the attribute of duration, which is compatible with another sense of *short* as an antonym of *long* (Fellbaum et al., 1993).

Polysemy, in the net of verbs, varies according to the frequency of the verb. WN, similar to linguistic ontologies, extracts data from dictionaries and imposes taxonomic relations on it. The most frequent verbs, e.g., *have*, *be*, *do*, and *make*, express several senses based on the semantics of their noun arguments, not because of the core semantic features of the verbs themselves. The difference between the senses of *have* in *have a car* and *have a disease* is based on the abstract and concrete noun arguments of *have*. Although separated in some dictionaries, the two senses are lumped in WN. WN splits verb senses which differ in one or more core components of the verb meaning. To elaborate, the sense of *beat*, which is synonymous to *strike* and *hit*, places the verb in the semantic field of "contact". The synset of *beat* and *flatten* highlights "change" as a core meaning component. Although it has contact as a core meaning aspect, the new component positions the verb in another semantic field (Fellbaum et al., 1993, WN 1.3).

3.1.2.2 CorLex Database

Fine granular splitting of polysemous senses in language resources is a challenge for WSD systems, which are supposed to choose the correct sense and assign it to each contextual word use. Therefore, several attempts have been made to modify sense separation in existing language resources and improve their usability in NLP tasks (Lenci, 2010). In addition to their fine granularity, senses in language resources are, in several cases, insignificant to domain-specific tasks. Domain-specific senses in WN are either missing or unrelated to domain-specific terms (Buitelaar, 2010). Despite the absence of a clear model for sense separation in WN, the database

is frequently used by ontological models that either use its current form or modify it for specific purposes.

CorLex is a top-level ontology and a semantic database of 40,000 WN-based nouns. It aims at describing what is thought to be the core vocabulary of the language. Related senses of a polysemous word are believed to be peripherally generated in the context of use. Therefore, the fine-grained sense distinctions in WN are generally lumped in CorLex's ontological levels. CorLex replaces WN's 60,000 (synset) tags of nouns with a smaller set of 126 semantic types (Buitelaar, 1998). This project considers WN a semi-formalized dictionary that categorizes words into classes. WN's several hierarchical levels and relatively comprehensive coverage of concepts are strongly associated with the ambiguity of its words. If the lexicon tackled only the highest ontological levels OBJECT and EVENT, almost all the classical cases of ambiguity (polysemes and homonyms) would be resolved. The two meanings of *bank*, for instance, will be lumped under OBJECT (Buitelaar, 2010).

The CorLex model is theoretically motivated by systematic or regular polysemy. Providing classes of systematic polysemy in a language facilitates the process of splitting and lumping senses consistently. Nouns denoting ANIMALS, for instance, can systematically refer to the animal, its meat, skin or fur and nouns referring to CONTAINERS have the related senses of the container itself and its contents (Atkins and Rundell, 2008). CoreLex hypothesizes that the close and systematic relation between such senses does not always require disambiguation. WSD systems should attempt to place the word into its systematic polysemous class instead of determining which of its closely related senses is used in the target text (Buitelaar, 2010).

The creation of CorLex followed three steps. The first step is aimed at the reduction of WN senses into a formal set of basic types. *Book* as a noun had seven senses in WN at the time of creating CorLex. Tracing the higher hierarchical levels of *book* synsets showed their shared hypernyms. They either descend from ARTIFACT or COMMUNICATION. Accordingly, the seven senses are reduced into two semantic types. *Book* in CorLex is represented in a way that reflects its meaning as the content being communicated (i.e., COMMUNICATION) and the medium of communication (i.e., ARTIFACT). The second step is the organization of words into systematic polysemous classes based on the distribution of their basic types. The basic types resulting from the first step are clustered for each noun to identify the systematicity and idiosyncrasy of polysemous words. Unlike WN, the distinction between homonymy and

polysemy and the consideration of regularities between senses are central to CorLex. The classes are manually explored to remove classes of homonyms, such as the class of "act anm art". It combines the semantic types of ACT, ANIMAL and ARTIFACT STATE and includes nouns such as *bat*, *drill* and *hobby*. In Contrast, the "art atr sub" reveals a type of systematic polysemy that is not predicted by the literature but is intuitive. Nouns in this class, such as *fiber* and *chalk*, systematically refer to ARTIFACT, the SUBSTANCE it is made of and the ATTRIBUTE. The third step is the use of underspecified semantic type definition to represent the systematic polysemous classes. ANP, for instance, is underspecified semantic type that refers to nouns denoting ANIMALS of PSYCHOLOGICAL or conceptual nature (Buitelaar, 1998). This method achieves an effective WSD at higher ontological levels and still allows further filtration or specification of senses if needed (Buitelaar, 2010). The system would assign ANP to a particular use of *monster*.n, which is sufficient for several applications. The option to specify the semantic type ANIMAL or PSYCHOLOGICAL FEATURE will be available.

Also, Ide and Wilks (2007) argued that the fine-granular taxonomy of senses in the lexicographic practice might be suitable for human users. However, performing WSD tasks requires a general level of taxonomy that corresponds to the upper-level distinctions between homographs. Although this is a theoretically and lexicographically ineffective method, it would retrieve almost 100% success of several NLP tasks. They claim that the lexicographic vision is so idealized, whereas the more practical view would stick to the differences that humans and machines can "reliably" tell.

3.1.2.3 OntoNotes and Omega

OntoNotes aims at annotating multilingual corpora of English, Chinese and Arabic texts. It relies on various language resources to derive information, such as predicate structure, ontology and word senses. TreeBank and PropBank provide predicate structure information, WN contains word senses, and Omega hosts the required ontological knowledge. The project starts with the annotation of the most frequent 700 verbs, which are supposed to be the most polysemous. OntoNotes supports the argument for lumping the fine-grained senses in WN. To that end, human annotators rely on the subcategorization frames of words and the semantic classes of their arguments. First, annotators divide the WN senses of a verb into groups, define the general sense expressed by each group, and state the criteria for such grouping. *Drive* in WN has 22 senses which can be grouped into four sets. The first group is defined as "operating or traveling via a

vehicle" and is syntactically realized through two patterns "NP (agent) drive NP" and "NP drive PP". This group lumps 7 WN senses of *drive*. The proposed sets of senses are then checked against the Omega ontology with the aim of enhancing meaning representation by linking senses to an ontology (Hovy et al., 2006).

The Omega ontology itself is the result of merging the hierarchical levels of WN with the Mikrokosmos ontology. A New Upper Model (NUM) is introduced to allow this merge without sacrificing specific concepts in the lower ontological levels in WN. NUM has, at its higher levels, a number of mutually exclusive features. Summum Genus branches into an intangible object and a tangible object, which also branches into mutually exclusive nodes such as a tangible volitional object and a tangible non-volitional object (Philpot et al., 2003). In the newer version of the ontology, Omega 5, senses are manually assigned to polysemous nouns and verbs by expert lexicographers. Omega, WN, corpora and dictionaries are sources for the manually-prepared list of senses. In this regard, the proposed NUM features serve as effective differentiae of senses (Philpot et al., 2005)

WN data is represented in a new way to serve the purposes of Omega. Each synset is represented as a concept and attached to its decomposed gloss (decomposed into definitions, usages and examples). Concepts were renamed to maintain their uniqueness and distinguish each concept from other related and unrelated ones. Natural identifiers cannot fulfil these purposes. A natural identifier of a concept is the most frequently used word to refer to that concept. *Bank,* for instance, is a natural identifier of related and unrelated concepts (e.g., slopping land next to a river; slope in the turn of a track; financial institution accepting deposits and to conduct banking transaction with). Omega uses a hybrid approach that generates a set of candidate names for each concept based on eight methods. A concept can be named by the direct choice of a word from the synset words, the reference to its immediate parent, subconcept or indirect ancestor, among others. The algorithm used in the Omega project suggests reasonable names for the concepts, unlike the arbitrary results numerically generated by other algorithms. However, the quality of naming is incomparable to that offered by humans. It still resolves the ambiguities emanating from the use of natural identifiers (Philpot et al., 2003).

Chapter Four – Cognitive Linguistic Approaches to Meaning

Whereas Ostermann (2015) called for the systematic application of cognitive semantic theories in lexicographic practice, cognitive approaches to meaning appeared tens of years before her. This chapter explores meaning from a cognitive linguistic perspective. It mainly focuses on the theoretical contributions of Fillmore, Langacker, and Lakoff. The implementation of Fillmore's Frame Semantics in the FN database will be discussed in detail because the database will be used in the three experiments that will be reported in the coming chapters.

4.1 Meaning and Conceptualization

The influence of the empirical findings in cognitive psychology has been remarkable on the cognitive linguistic views on meaning. The ideas of fuzzy categories and schemas have been borrowed from cognitive psychology and became dominant in cognitive linguistics, which rejected the arguments for the autonomy of the language system. Rosch (1978) argued that human categorization has two principles. First, the principle of cognitive economy states that category systems provide the maximum amount of information through the least cognitive effort. The second principle is the perceived world structure. It refutes the hypothesis that the world is a set of arbitrary attributes. Categories should reflect the structure of the perceived world as accurately as possible. Categories do not have clear boundaries that separate them from each other. They, rather, have members and a prototype. In this regard, prototypes are the most apparent reflections of the "structure redundancy" of the category and the most salient examples of category membership.

Parallel to the influence of categorization was the influence of schemas. Bartlett's (1932) work on schemas and structured human knowledge was revived in the 1980s and had a great impact on cognitive linguistic theories. Bartlett (1932) explained that a schema is "an active organization of past reactions or experiences". Understanding, responding and remembering are all associated with constructing the elements of a schema in an organized way as a "unitary mass". Several scholars in the 1970s and 1980s, such as Piaget (1976) and Bruner (1980), used the concept of schemas in their theories of development and learning. Schemas were used to refer to any mental representation, and schemas derived from bodily experience (at the early childhood stage) have been given particular importance. The arguments that human knowledge (including linguistic knowledge) is derived from sensory-motor experiences have been dominant.

4.1.1 Conceptual Structures Underlying Meaning

4.1.1.1 Fillmore's Semantically Motivated Grammar

Before the spread of the cognitive revolution in the field of linguistics, Fillmore (1968a) attempted to develop a theoretical framework that could capture the similarities between words such as *hit* and *touch* and, at the same time, account for the different syntactic occurrences of verbs like *open* given their similarities in meaning. Motivated by unveiling language universals (not conceptual structures), Fillmore's Case Grammar presented the following case feature, which successfully covers the syntactic variations of ergative verbs such as *open, increase* and *break.*

+[_____ O (I) (A)]

The abovementioned case feature showed that a group of verbs must have at least an Objective case (O) regardless of its syntactic position in the subject or the object position. The Objective case refers to the thing that is opened, increased or broken, for instance. There are two optional cases which are the Instrument (I) which is used to cause the action, and the Agentive (A), which is either the volitional performer or the non-volitional force behind the action (Fillmore, 1968a). Such differences that fall within the intersection between semantics and syntax are currently reflected in dictionary entries as grammatical labels like "transitive and intransitive".

That was a considerable step towards the development of a theory of grammar that was motivated by semantics. The most significant steps that led to the foundation of the Frame Semantics theory appeared almost ten years later. Fillmore (1975, 1977) was inspired by Bruner's theory of cognitive development from bodily interaction and visual representation to symbolic language use and Rosch's theory of human categorization. Frames, in cognitive psychology, denote "schemata or frameworks of concepts or terms" that form a coherent system and "impose structure on some aspects of human experience". However, the frame-scene model of meaning formulated a linguistic-oriented definition of frames, which was different from the term's original use in cognitive psychology. Fillmore's early use of frame referred to "any system of linguistic choices", which could be a collection of words or a choice of grammatical rules associated with particular scenes (i.e., visual scenes, common cultural scenarios, interpersonal transactions and experiences). The concept of scenes was relevant to the paralinguistic schematic experience. This distinction was jettisoned at later stages of the theory. The description of *write*.v in this model

provided the following information. The frame of *write* is associated with a scene whose entities are "Writer", "Implement", "Surface" and "Product". Given that the prototypical Product of writing has to be linguistic, *write*.v was associated with the frame of LANGUAGE. In addition, the scene of linguistic communication is added to the action scene of writing when a sentence such as *I wrote a letter* is uttered. The linguistic choice of *letter* as the product of writing is associated with a scene in which a "Sender" communicates a "Message" with an "Addressee". Fillmore explained that exposure to a linguistic element that is associated with a certain scene typically activates this scene in the mind of the speaker. Similarly, exposure to this scene would recall the word(s) that are associated with this scene (Fillmore, 1977a).

Scenes and frames can address the category-boundary challenge. Participants should not be asked about their judgments regarding the belonging of an item to a category. They should be asked about their willingness to extend the frame that is already associated with a prototypic scene to cover another scene relevant to the word under exploration or to create a new frame, such as a widow for a woman who murdered her husband (Fillmore, 1977b).

In 1985, Fillmore confirmed the significance of interpretive frames to the theory of meaning and differentiated between language-dependent and language-independent frames. Language-dependent frames make a considerable contribution to meaning, despite the dominance of language-independent frames. To elaborate, the description of containers of "soap flakes" as *large* in American supermarkets referred to the smallest size of containers, while *jumbo, economy* and *family size* are used to describe larger sizes. This association between *large* and the smallest size is not relevant to a non-linguistic frame, i.e., it does not reflect world knowledge about size and proportions. It is, instead, created by language to reflect this specific situation. Frames were still able to categorize words according to their shared underlying schema. At the same time, frames could assign separate frames to semantically related words such *skip, hop* and *leap*, which reflect different schemas of pedal locomotion.

4.1.1.2 Fillmore's Frame Semantics

Fillmore, accompanied by the prominent lexicographer Atkins, integrated the lexico-syntactic description of words, the cognitive concept of frames and the innovations of corpus analysis in a lexicographic-oriented study of the lexeme RISK. After Fillmore proposed a frame-based lexicon, linguists realized the potential of adopting a frame-based approach to dictionary-making

(Petruck, 2011). The new proposal centralized the role of conceptual frames in grouping words (Ackerman, Kay and O'Connor, 2014).

Fillmore and Atkins (1992) aimed at making all information that speakers have about their language accessible to users in a frame-based dictionary. The new perspective on the lexicon maintained that understanding the meaning of a word depends on knowledge about the frame (i.e., cognitive structure), which motivates the concepts encoded by this word. The new lexicographic proposal included information about (a) relations between meaning and lexico-syntactic patterns; (b) relations among words evoking the same frame; (c) relations among polysemous word senses, and (d) relations between the frame of the target word and any relevant frames.

According to FS, a frame is a schematic conceptual structure representing a situation with typical participants or frame elements. The existence of a lexical item stipulates its correspondence to a frame (Sambre, 2010). Electronic dictionaries should include frame information in the definition of words and use frame elements to describe meaning. Although advanced users may not be interested in the description of the frame, it is essential to arrive at the definition of the target word. Frame information contributes equally to decoding and encoding purposes (Fillmore, 2003).

Fillmore and Atkins (1992) conducted a comprehensive corpus-based analysis of the lemma RISK. The study relied on hypotheses about the association of RISK with two subframes, namely, RISK_TAKING and RISK_RUNNING, and the existence of categories such as Victim, Valued Object, Harm and Actor. Fillmore and Atkins used "categories" to designate a concept similar to that of semantic roles and cases. The new term was introduced to suggest the frame-specificity of the roles played by the phrases surrounding the target word, as opposed to the previous statements about the universality of cases and semantic roles.

(5) (a) He *risked* a trip into the jungle

(b) He *risked* his inheritance

(c) The board was *risking* a liquidity crisis

The frame-based analysis of *risk*.v linked the different senses of the verb to each other and linked the word's meaning to its semantic and syntactic valence. *Risk* in 5(a) denotes the performance of an act which causes a possibility of harm to oneself, and it expresses a relation between an Actor

(i.e., *he*) and a Deed (i.e., *a trip into the jungle*). This relation includes two valence patterns, displayed in (V). Deed can be syntactically realized as a noun phrase in the direct object position or as a gerundial object. In this case, *chance, hazard* and *venture* are relevant words to *risk*. Example 5(b) displays another sense of *risk* as a verb. It refers to acting in a way that puts something in danger and therefore expresses a relation between an Actor (i.e., *he*) and a Valued Object (i.e., *his inheritance*). *Jeopardize, endanger* and *imperil* are associated with this sense of *risk.* The NP complement of the verb *risk* can correspond to a third category. Example 5(c) expresses a relation between an Actor, which is *the board*, and Harm, which is *a liquidity crisis.* Harm is a "potential unwelcome outcome", and it can be syntactically expressed as a nominal object or clausal complement, among others. The use of these "categories", which are parts of the frame structure in the definitions of *risk*.v, highlights the similarities and differences between its senses (Fillmore and Atkins, 1992).

(V) Risk ^{D}{NP}

Risk ^{D}{Gerund}

4.1.1.3 Frame Semantics in the FN Database

The process of lexicography in FN consists of four steps. First, the vanguard, at the preparation stage, writes an initial description of the frame, proposes the semantic components of the frame, suggests a working label for each element and creates a list of the words which are expected to belong to the frame. In the second stage, the manually-recorded data by the vanguard serves as input to the automatic sub-corpus extraction of the sentences containing the lemmas of the target words. Third, annotators select canonical examples, and novel uses for each lexical unit, perform semantic role labeling and describe the phrase type and the grammatical function of each frame element. The three steps can be situated within the traditional analysis process. The final step, which is performed by the rearguard, corresponding to the traditional synthesis stage, includes a review of the information recorded by the vanguard and the annotator and writing the final entry for the lemma, the definition of the frame and the description of its frame elements (Fillmore, Petruck, Ruppenhofer and Wright, 2003; Baker, Fillmore and Lowe, 1998).

The FN team was committed to using frames in explaining the meaning of words and describing the semantic valence of words. It was also committed to drawing general conclusions about word usage depending on corpus evidence. At the macrostructure level, FN was designed to be a

human-browsable and machine-readable database. Therefore, the storage and encoding of the information in the databases were different from its representation in the user interface. Since the frame-based analysis proceeded frame by frame, not lemma by lemma, FN created separate entries to frames and lexical units. The lexicon part of FN recorded information about lexical units, i.e., a pairing of a word and a single sense. For each lexical unit, FN registered information about parts of speech, tabulated valence patterns and links to their annotated representative examples in the "lexicon". The frame database recorded the frames, their descriptions, frame elements (i.e., the last term used to encode frame-specific semantic roles) and relations to other frames in the Frame Database. The last part of the project included sentences in which the target word appears, and the frame elements are realized according to the stored valence patterns. The frame of DRIVING, for instance, referred to a situation in which a "Driver" is essentially initiating and controlling the movement of a "Vehicle". In addition, "Rider" and "Cargo" are other frame elements in DRIVING. The frame inherits the structure of the general TRANSPORTATION frame. The following annotated sentences displayed two patterns of *drive*.v (Fillmore, Wooters and Baker, 2001; Baker, Fillmore and Lowe, 1998).

(6) (a) [$_{D}$ Kate] *drove* [$_{P}$ home] in a stupor

(b) Now [$_{D}$ Van Cheele] was *driving* [$_{R}$ his guest] [$_{P}$ back to the station]

At first, the FN team planned to record statistical information about the frequency of sense-pattern relations. However, the statistical information provided in FN to date is the number of syntactic realizations of each frame element and the numbers of each valence pattern as recorded in the annotated examples. The annotation schema used in the current version of FN is three-layered: Frame Element (FE), Phrase Type (PT) and Grammatical Function (GF).

4.1.1.4 Langacker's Conceptually-motivated Grammar

Langacker (1999, 2017) argued that grammar symbolizes meaning, and meaning is a type of conceptualization. Therefore, a cognitive account of grammar has to be rooted in a conceptual view of meaning. Conceptualizations include any mental experience, whether abstract or body-related. Similar to Fillmore's perspectives on relating any linguistic item to a paralinguistic conceptualization (i.e., frame), Langacker relates linguistic expressions to a broad scope of conceptualizations. Whereas the meaning of *trumpet* is associated with auditory and visual images, the meaning of *walk* essentially includes sensory motor experience. Whereas Fillmore

calls the conceptual structures underlying meaning "frames", Langacker (2017) refers to them as the "conceptual base". The conceptual base of a linguistic item comes from a set of simple or complex cognitive domains. Basic cognitive domains such as space, time, vision, and smell cannot be decomposed into sub-domains or less complex domains, whereas complex domains such as emotions, communication or relations are non-basic. The meaning of any linguistic item typically involves a set of domains. The encyclopedic (linguistic and non-linguistic) knowledge associated with a word is structured in the mind in terms of central and peripheral information/domains which overlap and intersect. The use of a word in a sentence activates part of this knowledge. The conceptual base of *glass*, for instance, involves multiple cognitive domains such as space, size, material, shape, function and cost. Similarly, the conceptual base of *dog* includes a set of domains like animals, sound, movement, shape and size. This matrix of domains is not arbitrarily stored in the mind. It represents structured knowledge accessed or activated differently based on the context. The centrality of a domain to the meaning of an expression controls its degree of activation. Whereas the function (of containing liquids) is central to *glass*, cost or washing of *glass* are less central and, accordingly, less likely to be activated.

As implemented in the FN database, the current version of FS refers to a similar point. Looking up *glass* in the FN database shows that the sense intended by Langacker activates the frame of CONTAINERS. The only core FE in this frame is the container itself (the first function identified by Langacker). The method of "construction", "material", and "use", among others, are included as non-core FEs. Although FEs are not equivalent to cognitive domains, both refer to core and peripheral structured human knowledge associated with the meaning of a frame and, or, a linguistic unit.

Relevantly, Lakoff (1987) developed his ideas about the structure of human knowledge in Idealized Cognitive Models (ICMs). His proposal intersects with Fillmore's FS, Langacker's Cognitive Grammar and Johnson's and Lakoff's CMT. He acknowledged the complexity of the conceptual structure of an ICM. He clarifies how several models cluster to give rise to basic concepts such as *mother* (e.g., birth model, genetic model, marital model). However, some real-life situations do not comply with one or more of these models (e.g., surrogate mother, adoptive mother). Hence, one model seems to be more central or primary than other models, and this is

usually reflected in dictionary definitions. However, dictionary editors may differ in their choice of the primary model.

4.1.2 Conceptualization and Construal

The cognitive approaches to meaning revealed similarities among superficially unrelated words, explained the conceptual links between grammatically related words/structures and unveiled the similarities and differences between semantically related words. Fillmore's approach was frame-based, not word-based. Therefore, it missed several points highlighted in Langacker's approach. However, the two approaches had some points in common.

4.1.2.1 Fillmore's Frame-to-Frame Relations

Ruppenhofer, Baker and Fillmore (2002) aimed to improve the understanding of frames and add robustness to the FN database by adding frame-to-frame relations. Ruppenhofer et al. (2016) explained the various conceptual relations that hold between the frames in the current version of the database. Frame relations reveal the shared meaning aspects between the LUs that evoke different frames. FN includes 13 types of frame-to-frame relations. The first pair of relations is "inherits from" and "is inherited by". At the frame level, inheritance is a relation between a general parent frame and a more specific one. For instance, MOTION is inherited by SELF_MOTION, which is inherited by TRAVEL. This hierarchical relation explains how *move, come, circle* (MOTION), *walk, run, march* (SELF_MOTION) and *travel, journey* and *tour* (TRAVEL) share the same conceptual structure but specify it differently. The first group evokes the MOTION frame, which has "Source", "Path", "Goal" or "Area" and "Theme" as core FEs. The "Theme" has to be any physical object (*she, boat, animal* appear as fillers of "Theme" in this frame). The SELF_MOTION frame shares, but specifies, this structure. It replaces "Theme" with "Self_mover" which has to be sentient. This excludes the inanimate fillers allowed in MOTION. The third group, which activates the frame of TRAVEL, adds another core FE to specify the frame: "Mode of transportation".

The second pair of relations is "Perspective on" and "is prespectivized in". The "perspective" relation holds among at least three frames. A general parent frame expresses a scenario that can be viewed from two perspectives (frames). "Perspective on" highlights how *give, donate* and *hand over*, are related to *receive* and *accept*. Whereas the first group of verbs evokes the frame of GIVING, the second group evokes RECEIVING. Both frames adopt a "perspective on" the

TRANSFER frame. TRANSFER refers to a situation in which the possession of a "Theme" is transferred from a "Donor" to a "Recipient". Whereas GIVING adopts the perspective of the "Donor", RECEIVING adopts the perspective of the "Recipient".

Third, "Uses" and "Is used by" are relations between a general frame and a more specific one. The general frame must be in the background knowledge of a user to understand another child frame. Unlike the inheritance relation, the child frame is not a specific case of the general frame. The frame of ROADWAYS, for instance, is evoked by LUs such as *sidewalk*, *street*, *highway* and *road*. Understanding ROADWAYS requires knowledge about MOTION.

Some frames express complex scenarios that usually occur in a specific order. Four relations are used in FN to reflect such complexity (i.e., the pairs "has sub-frames" – "is a sub-frame of" and "precedes" – "is proceeded by". The CRIME_SCENARIO, for instance, is a complex frame. It has the subframes COMMITTING_CRIME, CRIMINAL_INVESTIGATION and TRIAL. COMMITTING_CRIME has to precede the CRIMINAL_INVESTIGATION, which also precedes TRIAL. The three frames typically occur in the previous chronological order.

The last pair of relations between frames is "Causative of" and "inchoative of". FN identifies this relation between several frames. CAUSE_TO_START, CAUSE_CHANGE and CHANGE_POSTURE are causative of PROCESS_START, UNDERGO_CHANGE and POSTURE, respectively. Finally, "See also" is a newly added relation, which is a link between a set of frames forming a group. Seeing the definition of one of them is enhanced by seeing the definitions of other frames in the group.

4.1.2.2 Langacker's Cognitive Abilities

Langacker (2017) stated that the meaning of a linguistic expression is not only the conceptual base of this expression but also the ways in which this base is construed. Meaning is the conceptual basis and the construal. Three dimensions of construal are frequently stressed in Langacker's work. First, the level of specificity versus schematicity can result in various utterances describing the same situation. At the most schematic level, "thing" can replace nouns and "do" can replace verbs. Adding adjectives and adverbs increases the specificity of a sentence. Whereas *something happened* and *a dirty poodle entered the kitchen* can be used to describe the same situation, they reveal different degrees of specificity. Fillmore's inheritance relation somehow reflects these word and sentential levels of specificity. However, this requires frame

annotation for each LU. *Happen* evokes the EVENT frame at the top of the frames hierarchy, and *enter* would inevitably evoke a more specific frame. Also, *thing* evokes the top-level frame ENTITY, and any other noun would activate a more specific frame. The two hierarchies EVENT → EVENTIVE_AFFECTING→ARRIVING and ENTITY→ BIOLOGICAL_ENTITY → ANIMALS reveal the levels of specificity in the subject and the verb of the abovementioned sentences. The specificity emanating from the adjectives or the adverbs is generally reflected in the instantiation of non-core FEs.

The second dimension of construal is prominence, which has different types. Profiling (kind of prominence) refers to what the expression focuses on, given its conceptual base. Whereas verbs profile processes, nouns profile things. In pairs like *watch-watcher, love-lover*, *admire-admirer*, *perceive-perceiver*, the verb and the noun share the same conceptual base. For instance, the conceptual base of perceive-perceiver is a phenomenon and a person who perceives it. The verb profiles the entire relationship, and the noun profiles or focuses only on the person who perceives the phenomenon. A second case of profiling is relevant to part-whole relations. *Teeth, tongue* and *gum* share the conceptual base of the mouth, but they profile different parts. Also, understanding the relation between a parent and the offspring is the conceptual base of both *parent* and *child*. However, *parent* profiles the older member of the relation, whereas *child* profiles the younger one. Verbs like *give-receive* have the same conceptual base in which a giver transfers an object to the receiver. *Give* profiles the relation between the giver and the object, but *receive* focuses on the object and the receiver.

Only the last profiling case is reflected in the FN database (perspective-on relation). The differences between the verb and the doer (e.g., *teach-teacher*) are usually present in the valence description of the two words without reference to the conceptual differences motivating them. Also, the different profiling of the participants in relationships is missed from the FN database.

The organization of the trajector and the landmark is another way of changing the prominence. Langacker names the two entities in a relation trajector (which is the primary focus) and a landmark (which is of secondary prominence). This is most salient in pairs like *below-above* or *before-after*. Whereas *X happened after Y* and *Y happened before X* describe the same situation, the organization of the trajector and the landmark are different. *After* gives the focal prominence to X but *before* focuses the attention on Y. Such differences are absent from dictionaries, but they are incorporated in FN. Although such pairs are usually placed in the same frame, the core FEs

are sufficient to reflect the prominence of the participants (usually Figure and Ground or Event and Landmark Event).

Langacker (1990) explained the third aspect of construal, which is perspective. Pairs like *above-below* and *before-after* describe the same situation differently by changing the focus from one participant to another. However, they do not denote a change in the described situation itself, given the verticality of the human body and chronological time sequence. Perspective, however, refers to differences in the perceived situation itself. According to one vantage point, a situation may be described as *x is behind y*, but from another vantage point, the same situation may be described as *x is in front of y*. Therefore, prepositions like *above* profile a two-entity relation (trajector and landmark), but prepositions such as *in front of* has a third element in the relation (i.e., vantage point). In the FN database, prepositions which essentially profile a vantage point are grouped in the NON_GRADABLE_PROXIMITY frame. This frame is evoked by *in front of, in the back of, right, left,* among others.

The fourth aspect of construal is the scope. Langacker differentiates the overall scope, which motivates the conceptual base of an expression, from the immediate scope, which is activated based on the context. To elaborate, the overall scope of *glass* includes space, containment, size, shape, use, material, cost, drinking, beverages and more. *Crystal glass*, *wine glasses* and *pint glass* activate different immediate scopes. *Crystal glass* essentially activates the material and probably the cost. *Wine glass* has use and content as an essential part of its immediate scope. The immediate scope of *pint glass* activates size and measures. Langacker highlighted a noticeable grammatical pattern in forming noun-noun compounds such as *fingertip*, *eyelashes* and *eyelid*. The first noun is the immediate scope of the second noun.

Dictionary definitions usually refer to the immediate scope of a word. However, the overall scope is rarely made available. The new thesaurus features in online dictionaries such as Cambridge's SMART vocabulary and Oxford's Topics offer part of the overall scope of word senses. To exemplify, OALD assigns "building" as the topic relevant to the sense of *glass* as a substance and assigns "cooking and eating" to the sense of *glass* as a container.

4.2 Polysemy and Categorization

Lyons (1977) described polysemy as a case in which a lexeme has multiple meanings, and the relationship among them is historically and, or, semantically present. Apresjan (1974) defined the

semantic relatedness between word senses as shared components in their definitions. He further classified polysemy into regular and irregular types. Whereas regular polysemy exhibit patterns that can be detected across a group of words, irregular polysemy does not reflect such patterns. Regular polysemy includes metonymic patterns like “animal for meat”, “container for content”, and “tree for its wood”. The cognitive linguistic approach to polysemy views it as a type of categorization. The different senses of a polysemous lexeme form a category whose senses (members) are clustered around a prototypical sense. They display differences from each other and bear degrees of family resemblance. The more components they share with the prototypical sense, the more central they are to the category (Halas, 2016; Lakoff, 1987).

Langacker (2017) explained that the polysemous senses of a lexical item form a network with categorizing relationships linking them. He referred to two main types of categorization relations: elaboration and extension. Elaboration denotes a relation between a schematic sense A (e.g. the schematic sense of *ring* as a circular entity) and a more specific sense B (e.g. the sense of *ring* as a circular mark or a circular object). The elaboration relation links the general sense of *ring* as a circular object and the more specific sense as a circular piece of jewellery. Extension holds between a prototypical central sense and a peripheral one. Metaphoric extensions show similarity between the central and the peripheral senses, and metonymic extensions show an association between the senses.

Lakoff (1987) explained that the different senses of a word usually involve different ICMs. The birth model is essential to the metaphoric extension of *mother* in *necessity is the mother of need*. However, the genealogical model gives rise to the metaphoric extension of *mother* in the linguistic syntactic trees.

Halas (2016) introduced another prototype-based model for addressing polysemous words in dictionaries. The model clarifies the hierarchical organization of polysemous words, family resemblance among related senses and the motivation for deriving peripheral senses from core ones. The model helps lexicographers modify dictionary entries to systematically account for senses differentiation, definition and order. The first step lexicographers should take is identifying the prototype as the basic sense. The examination of the corpus citations of a lexeme should lead to the detection of its semantic base. Second, identifying primary and secondary senses should be based on a motivational analysis of sense derivations. This analysis includes detecting the derivational path from the center to the periphery and displaying the derivation

mechanism. Lexicographers make decisions to establish the superordinate sense that is general enough to include sub-senses and split subsenses according to their distinctions.

According to another prototype-based approach, polysemous words are clustered around more than a single prototype category. Polysemy can be intercategorial (i.e., appear across category boundaries) and intracategorical (arise within a single category). To elaborate, the prototypical sense of the verb *command* is "giving orders through a variety of mediums including voice, paper and emails". However, the meaning of *command* crosses the boundaries of this sense to two other categories relevant to perception (e.g., *the house commands a view of the lake*) or abstract manipulation (*the fortress commanded the valley*). The polysemy of c*ommand* is arguably clustered around several radially-related categories, each with its members (Lewandowska-Tomaszczyk, 2007).

Homonyms, however, are accidentally related and lack etymological and semantic connections. Therefore, homonymy is described as arbitrary relation between two distinct lexemes. The word form *found* corresponds to a meaning denoting the past tense of the verb *find* and to the base form of the verb *found*. Similarly, the word form *bank* corresponds to the meaning of a financial institute and a riverside. *Bank* represents a case of absolute homonymy. The two readings of the word are unrelated; the word forms of *bank* are shared between the two meanings and are grammatically equivalent. The ambiguity of *found,* however, emanates from partial homonymy relation. Not all word forms of *found* and *find* are shared between the two meanings, and the shared form is not grammatically the same in the two readings. *Found* is a transitive verb in the two meanings, but the tense of the verb differs, and accordingly, there is no grammatical equivalence. Therefore, modifying the grammatical context of *found* through the use of a singular subject or continuous tense, for instance, would resolve the ambiguity (Lyons, 1977).

Several counterarguments are made against the distinction between polysemy and homonymy based on etymology and semantic relatedness. Alekse (2017) argued against the theoretical distinction between polysemy and homonymy. The distinction based on etymology and semantic relatedness is arbitrary and not straightforward. Some word origins cannot be traced through history. It is unclear to which historical stage a lexicographer or a linguist should go back to find the origin of two word senses. It is also challenging to determine a word's core meaning and clarify how central and peripheral meanings descend from it.

Apresjan (1974) himself stated that polysemy and homonymy are not absolute, and they do not have clear boundaries. For instance, metaphorically motivated polysemy is close to homonymy, while the type motivated by metonymy is not. That is to say, the predictability of and motivation behind polysemy at the theoretical level do not help in telling apart cases of polysemy from that of homonymy. In addition, polysemy can be motivated by foreign influence and semantic borrowing.

Some counterarguments against the etymological polysemy-homonym distinction discussed the famous *bank* example. Tóth (2008) displayed how the literature proves the relations between the two senses of *bank* although they are not etymologically related. Although the two senses of *bank* descend from Italian and Scandinavian origins, bankers used to be available at the riverbank. That is to say, going to the *bank* as a "financial institution" entailed going to the *bank* as the riverside where bankers were. Second, in several cases, the knowledge of native language users about the etymology of words is absent. Hence, the etymological difference no longer has psychological relevance.

Another counterargument against the widely cited example of ambiguity *I went to the bank* followed a usage-based approach. Examining the corpus instances of the word showed that *bank* rarely causes ambiguity. The possibility of having two readings of *bank* in a single context was not present in 1,000 occurrences in the BNC. Interestingly, corpus citations showed instances that were not ambiguous but that corresponded to neither the financial institution meaning nor the riverbank. They include *blood bank*, *data bank* and *seed bank* (Hanks, 2000). Moreover, examples such as *solid bank, frozen bank account, sand deposits* combine two features from the two supposedly homonymous senses of *bank*. The conceptual metaphor MONEY IS LIQUID motivates the meaning of some of these combinations.

4.2.1 Shared Conceptual Structures

On the one hand, Fillmore's approach detects polysemy with reference to the semantic arguments of a word and their syntactic manifestations in a sentence. Although the current version of FS in the FN database does not distinguish polysemy from homonymy, Fillmore's early work dealt with polysemous verbs extensively (e.g., *risk*). At the cognitive level, the different senses of a word generally evoke separate frames and are, accordingly, split into various entries according to their respective frames (Fillmore, Johnson and Petruck, 2003). At the linguistic level, the

annotation of corpus instances presents evidence of the existence of different senses. During the annotation process, lexicographers apply several steps to split or lump the different uses of a word. If a word occurs with different numbers and, or, types of arguments, two or more senses should be separated. This criterion systematically separates, for instance, the transitive and intransitive senses of ergative verbs. However, in this case, the shared conceptual structure between the two senses is reflected in the "causative of" relation between their frames. Also, two or more senses of a word are detected if the semantic relation between the word and its arguments differ. *Write.v* may label its arguments "Speaker" and "Message" in the STATEMENT frame or "Author" and "Text" in the TEXT_CREATION frame. Accordingly, two senses are realized in this case. The conceptual link between the two frames is their "Use" relation to the frame of COMMUNCATION (Ruppenhofer et al., 2016; the FN database, 2021).

However, FS does not categorize a word's (polysemous) senses according to their prototypicality, metaphoric or metonymic extensions. Therefore, the relation between the senses of *glass* as a substance and as a container is missing, and no conceptual links between their respective frames are identified in FN. Similarly, the senses of *employ* as hiring a person and its extension as using an instrument are dissociated in the database.

On the other hand, Langacker's approach succeeds in categorizing the polysemous senses of a word even if clear metaphoric and metonymic extensions are absent. The argument is that meaning is a construal imposed on conceptual content. The types of categorizing relations can be used to construct a network of the polysemous senses of, at least, most words.

4.2.2 Metaphoric Extension

The connection between metaphors and polysemy in lexicographic practice dates back to Dr. Johnson's plan for a dictionary. He clarified that the metaphoric sense of *arrive* (i.e., reach a desired goal) is to be separated from its literal sense (i.e., reach a place) (in Fontenelle, 2008). It was then reiterated in the research on regular polysemy. After introducing the Conceptual Metaphor Theory, several meaning-related aspects changed; some had salient implications for lexicographic practice.

Johnson and Lakoff (1980) proposed that metaphors are ways of thinking, and they are frequently present in everyday language. Several conceptual metaphoric structures, such as ARGUMENT IS A WAR, underlie commonly used expressions like *claims are indefensible*, *demolish or win an*

argument. Regarding the use of CMT in sense separation (metaphoric and non-metaphoric senses), the Metaphor Identification Procedure, hereafter MIP, (Pragglejaz Group, 2007) effectively reduces the subjectivity of applying the CMP. It starts with reading the whole text to understand its general meaning and decompose it into lexical units. Then, the most significant step, which is applicable to lexicographic practice, is the analysis of the target lexical unit in its context. This step involves (a) determining its contextual meaning given its surrounding words; (b) deciding whether it has more basic (e.g., concrete, body-related, precise, or older) meaning in other contexts and (c) if the contextual reading of the target lexical unit differs from its basic meaning, the lexicographer or the scholar should decide whether the contextual meaning can be understood against the basic meaning or not. *Struggle* in *Sonia Gandhi has struggled to convince Indians…* is marked as an instance of a metaphoric use. Whereas the basic meaning of *struggle* denotes "the use of physical strength against someone or something", the contextual meaning refers to exerting effort and facing difficulty in achieving a goal. Contextual meaning is easily understood against the background of the basic meaning. Determining the basic meaning of function words, however, is more challenging than identifying the basic meaning of content words.

Siqueira et al. (2009) applied the MIP and added two more criteria that are helpful in the detection of metaphoric polysemy of lexical items in the *Dicionário de Direito Ambiental*, a terminological dictionary on environmental law. The first criterion is the productivity of the conceptual metaphor. The uses of *in* in *with no further fishing in 24 hours* and in *seas, rivers and lakes in the territory of a State* are marked as highly metaphoric and literal, respectively. In the latter, the literal sense associated only with the CONTAINER image schema is evident. However, *in 24 hours* combines the CONTAINER schema with the conceptual metaphor TIME IS A SPACE. Accordingly, the contextual meaning differs from the literal one and can be understood against it. The productivity of this metaphor is marked in commonly used expressions such as *my birthday is near*, *far from winter*, *deadline has been moved up*, among others. The second criterion is resisting literal paraphrasing. If the lexical unit is used metaphorically in a particular context, using a word that is synonymous with the target word's basic sense results in meaning distortion. The use of *high* in *preservation of natural ecosystems of high ecological relevance* is an instantiation of the metaphor IMPORTANCE IS SIZE. This metaphoric contextual sense of *high* (i.e., important) is replaceable with other lexical items from the same field of size, such as

little or *great*, without significant variation in the meaning. However, replacing *high* with a word that is synonymous with its literal sense would distort the contextual meaning.

Dai, Wu and Xu (2015) studied the impact of using conceptual metaphor information, in *Macmillan English-Chinese Dictionary for Advanced Learners* (MECD), on learning metaphoric collocations. At the cognitive level, they investigated the impact of explaining conceptual metaphoric information on the retention of metaphoric collocations. At the lexicographic level, they measured the time of lookup and retrieval of metaphoric collocations from MECD if compared to the *Oxford Advanced Learner's English-Chinese Dictionary* (OALECD), which lists metaphoric collocations according to their lexico-grammatical associations. Participants in the control group were provided with entries retrieved from OALECD. Therefore, the conceptual metaphoric information was neither available nor salient. The first target group was exposed to conceptual metaphoric information in the most salient way because conceptual metaphors were listed immediately after the headword, and minimally seven sentences demonstrated the conceptual metaphors. The information about conceptual metaphors was available but less salient to the second target group; it was placed after the metaphoric sense, and a single sentence exemplified it. The availability and salience of the cognitive information shortened the time of the lookup and retrieval of collocations and promoted the retention of metaphoric collocations.

Van der Meer (1999) promoted the transition from literal to metaphoric meanings in dictionary entries as the ideal semantic arrangement. The metaphoric approach to exploring sense representation enriches the awareness of dictionary users about meaning extension and transition from the literal to the figurative senses. Therefore, the ideal dictionary entry should centralize the literal meaning and crystalize the relevance of other possible meanings to it. In more recent studies, Dalpanagioti (2019) constructed a new dictionary entry for the verb *stagger* and ordered the senses following the conceptual metaphor approach. The edited entry clusters the meaning of *stagger* around two frames, namely, self_motion and cause_motion. The first cluster, for instance, starts with the literal meaning, i.e., "walk or move unsteadily as if you are going to fall over", followed by the figurative meaning "continue or carry on with great difficulty". The second figurative meaning is derived from the first one through the conceptual metaphor that MANNER OF ACTION IS MANNER OF MOTION.

Mheta (2017) utilized both CMT and a Prototype-based model to study metaphors in three terminological dictionaries in Shona. After the detection of conceptual metaphors in the

dictionaries, the Prototype-based model was used to describe the similarities between the concepts and identify the best example of each category. The three studied dictionaries were overwhelmed by concrete and abstract metaphors. Animistic metaphors were central to the category of concrete metaphors. The "best exemplar", or the prototype of animals, in the studied dictionaries was the human being. The body parts of the human being were typically used to describe musical terms. In this regard, the metonymic use of *head* was the basic category, while *seeing*, *breathing*, *talking* and *thinking* were the attributes. The *guitar*, for instance, was described as a human being that has a *head*, *neck* and *waist*. The *head* was considered the most important part because all the essential parts of the guitar are attached to it (e.g., the strings that produce the sounds and the pegs that adjust the tones). Moreover, *head* was included in all the metaphoric meanings of *importance.*

4.2.3 Metonymic Extension

X stands for Y is the usual formula used to indicate metonymy (Johnson and Lakoff, 1980). Lakoff (1987) explained metonymic relations in terms of ICMs. He argued that metonymic models are types of structured knowledge that allow a "stand for" relation between two elements in an ICM. The concept that is understood in terms of the other is associated with it or in a part-whole relation with it. Some metonymic extensions have been noticed by linguists and lexicographers as cases of regular polysemy (e.g., animal for its meat, container for contents).

Langacker (2017) added another explanatory dimension to the metonymic extension of meaning. The literal and metonymic senses of a word share a conceptual base, but they are profiled differently. For instance, the literal sense of *church* profiles the building itself, and the metonymic sense profiles the religious organization.

Categorizing the senses of *fresh* in OED provided a new taxonomy based on the similarity of senses, metonymical extensions and specialization. The senses "new or novel" and "recent and newly made", for instance, are split in the dictionary, but they are similar. In addition, the sub-senses grouped under "new, recent" are metonymically related to the sub-senses under "having the signs of newness", which is relevant to perishable food and water. This analysis offers a new representation of the senses and shows how the polysemous readings are related. Still, it does not present a new splitting or lumping strategy (Geeraerts, 2001).

Sweep (2010) adopted a frame-based approach to the study of logical metonymy. The study focused on one of the dominant logical metonymies, i.e., OBJECT FOR ACTION IN WHICH THE OBJECT IS INVOLVED. To elaborate, the semantics of verbs like *begin*, *finish* and *enjoy* is compatible with an activity or an event. However, these verbs may combine with a metonymical object typically involved in the event instead of mentioning the event itself. *Begin a book* is interpreted as 'begin the activity of reading or writing a book'. The study focused on 21 verbs divided into three groups, namely, aspectual (e.g., *begin, start, continue*), evaluative (e.g., *choose, enjoy, want*) and in-between (e.g., *attempt, resist, survive*). The frame-based analysis identified 13 frames triggered by these verbs. In addition, the metonymical object did not typically correspond to part of the frame evoked by the verb. It was part of a different frame.

Two conceptual structures (i.e, frames) are present in *Mary began the book*. The first frame is ACTIVITY_START, and it is evoked by *began*. It has "Agent" and "Activity" as core frame elements. Whereas "Agent" is realized by *Mary*, "Activity" is not present at the surface level. In addition, *the book* is not a filler of any frame element slot in ACTIVITY_START. Rather, it is part of the "Activity" that is not linguistically realized. It is a filler of the core frame element "Text" in the READING_ACTIVITY frame. Therefore, the previous example contains an embedded structure relevant to the implicit "Activity" frame element and another main conceptual structure evoked by the main verb in the sentence, i.e., *began.*

This analysis revealed the presence of another logical metonymical relation PARTICIPANT (AGENT) FOR AN ACTION which accounts for several sentences such as *Mary interrupted John*. *Interrupted* evokes the frame of INTERRUPTING_PROCESS in which "Agent" and "Process" are core frame elements. On the one hand, *Mary* realizes the frame elements of "Agent". On the other hand, *John* is relevant to the "Process", which is a process of talking or presenting. *John* is, therefore, the "Interlocutor" in the embedded conceptual structure of DISCUSSION.

Chapter Five – Comparing Sense Delineation in OALD, FN and WN

5.1 Methods and Data Collection

This chapter describes how the sense delineation systems in OALD, FN and WN are compared in this study. Given the unavailability of a direct method to test the delineation system, the study compares the realization of the three methods as reflected in the senses listed in the entries. The exploratory experiment aims at answering the following questions:

1. Do the different sense delineation methods used in OALD, FN, and WN influence the encoding and decoding performances of ESL learners?

2. Which of the three approaches is optimal for the encoding and decoding performances of ESL learners?

The following sections elaborate on the steps of conducting this classroom-based experiment. They also report and discuss the results of the experiment.

5.1.1 Selection of Words

The experimental test examined five words belonging to three parts of speech. *Appear*.v, *tell*.v, *level*.n, *development*.n and *fair*.a are the five chosen words. The criteria for including words in the study were as follows. First, each word must have at least five senses in the three language resources. This condition was challenging in FN's incomplete database. Therefore, the second criterion was having at least an initial entry (e.g., containing the definition and the frame evoked by the LU, even if the valence description was absent) for each word sense in FN. Otherwise, the word was excluded from the experiment. Third, words must share a comparable number of senses in the three language resources. Overall, the tested entries contain from six to ten senses. All entries must have at least one sense that is saliently shared among the three resources so that the test questions can be unified for the three entry types. Finally, a signpost must precede senses in OALD entries.

The target words in the entries were replaced by words that are not currently used in English in order to avoid the influence of previous exposure. Chan (2014) pointed to the effect of word familiarity on the dictionary consultation process. First, ESL learners share a false assumption that they should consult an MLD only when they encounter an unfamiliar word. Moreover, if they consult a dictionary to check a word they are familiar with, they become influenced by their

preconceptions about the word's use or meaning. They usually skip useful dictionary definitions, examples and grammatical patterns when they find the meaning or the use they are already familiar with. This usually results in unsuccessful dictionary consultations. Second, Dziemianko (2016) explained the importance of changing the target word in a lexicographic test with an unfamiliar word to prevent the participants from relying on their background knowledge. Two methods can be used to replace the target word. Researchers either coin a word that complies with the phonotactics of the target language or use a rare word that is very likely to be unknown by students. For the second option, the *Compendium of Lost Words* was recommended as a source of rare words. Lost words in the *Compendium of Lost Words* are not present on web pages or online dictionaries. Therefore, it is improbable for learners to encounter them. The *Compendium of Lost Words* includes more than 400 words from the OED that are not in current use in English.

Lew and Pajkowska (2007) and Ptasznik and Lew (2019) chose to keep the original target word in the test. They chose a test sentence that instantiates the less frequent sense of the word. This sense is usually positioned towards the end of the entry. This strategy attempts to reduce the influence of previous exposure and encourage students to read the entire entry (Lew and Pajkowska, 2007; Ptasznik and Lew, 2019). However, this strategy is not effective for the current study, given the multiple dissimilarities between the three language resources. There is a remote possibility of finding a less frequent sense delineated and presented in the same way in the three resources. Moreover, this strategy does not totally undermine the influence of previous exposure as an interfering variable in the experiment.

The current study followed Dziemianko's (2016) method and used the recommended resource to select substitutes for the target word. Substitutes belong to the same part of speech as the target words to avoid user confusion. *Appear*.v was replaced with *avunculize*.v, which appeared and died in 1662. *Addecimate,* which was in use from 1612 to 1755, substituted *tell*.v. *Aeipathy*.n replaced *level*.n. It first appeared in 1847 and was last used in 1853. *Assectation*.n used for the single year 1656 was substituted for *development*.n. Finally, *autexousious*.a, which appeared and disappeared in 1678, replaced *fair*.a.

5.1.2 Designing Dictionary Entries

The second step in the experiment is modifying the lexicographic entries of the three projects to concentrate only on sense description. The three projects have different macro, micro and access structures that should also be modified to reduce the variables and focus on the goal of the experiment. Modifying a dictionary entry for test purposes is a common practice in lexicographic studies. Ptasznik and Lew (2014) modified dictionary entries to display both signposts and menus in the same entry. Dziemianko (2016; 2017) changed the position of the correct senses and altered the presentation of signposts in multiple dictionary entries. Ptasznik and Lew (2019) also modified the number of senses in dictionary entries for test purposes.

The original entries in OALD contain morphological information (e.g., word forms), phonetic information (i.e., transcription of the word), grammatical information (e.g., transitivity of verbs, some syntactic patterns), syntagmatic relations (e.g., collocations), semantic relations (e.g., synonyms), signposts, example sentences and information about word level based on the Common European Framework of Reference for Languages (CEFR). The current study is only concerned with the results of the sense delineation approach adopted by the OALD lexicographers. Therefore, it shortens the OALD entry to word senses and signposts. It keeps the structure of presenting information in its traditional dictionary format because of its familiarity and accessibility to learners. OALD entries are the least subject to modifications. Figure 7 shows a sample of the original OALD entry and the modified version used in the test. Only the information inside the red square is kept in the modified entry, whereas the discarded information is struck through.

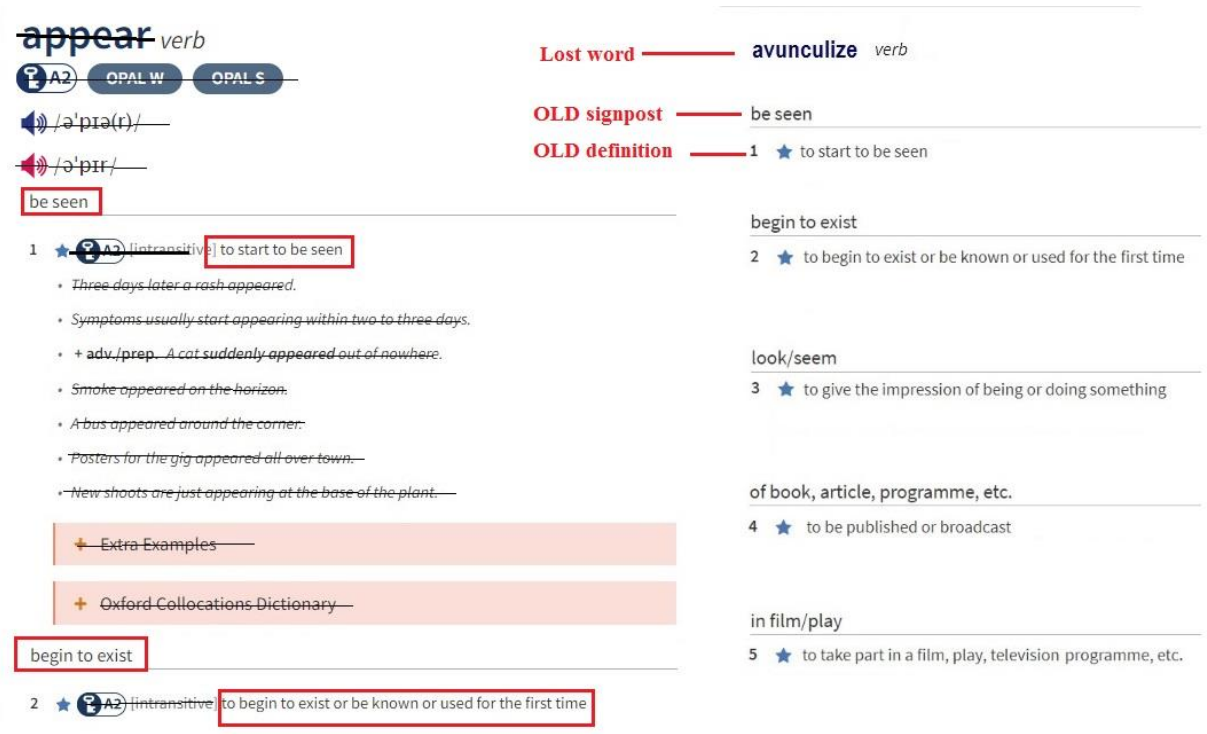

Figure 7. OALD original and modified entries of *appear*

The FN entries have been reconstructed to match the traditional access structure and microstructure of dictionaries. First, instead of having multiple entries for the different word senses, a single entry is used to record them all. Second, all the frame-based and syntactic information provided in the frame entries or in the LU entries is discarded from the modified entry. Similar to OALD entries, only the different senses (i.e., LUs) of the lemma are kept. Instead of signposts, frame names are used to add an organizational layer to the entries. The exact format is used to present FN senses. Definitions in FN are already cited from *Concise Oxford Dictionary* (COD) in most cases and are written by FN lexicographers in other cases. This information was removed from the modified entries, too.

WN entries are modified in the following way. The synonyms and the gloss are kept for each sense because WN does not provide complete definitions of the sense, and it records several fine-grained senses. Keeping only the gloss would result in considerably shorter and less informative entries if compared to the OALD and FN entries. In addition, the Semantic Concordance (SemCor) corpus is used to determine whether further information about the hypernyms, hyponyms or troponyms is frequent enough to include in the entry. Semcor is a sense-tagged

corpus that used WN as a lexicon of senses and definitions to annotate part of the Brown corpus (Landes et al., 1998). The current study makes use of the SemCor version hosted in Sketch Engine.

There was no feature that could be used as a guideword for WN senses. First, synonymy is the only relation that is consistently applied to all POS nets, and therefore, it seems to be the best candidate for guidewords. However, synonyms are not always available for each sense. For instance, three of the eight senses of *appear*.v do not have any synonyms recorded in WN. Moreover, sometimes WN records synonyms that are infrequently used in this sense and would, accordingly, perplex users for either the synonym's unfamiliarity or familiarity in another sense. For instance, *euphony*.n is the only synonym of the second sense of *music*.n. The synonym would be more challenging to learners than the target word itself, and it would not serve the purpose of the guideword.

The same applies to the use of WN's hypernyms as guidewords. In many cases, the hypernym is too general to provide helpful information to learners. For instance, *process*.n is recorded as a hypernym of *development*.n and *be*.v is recorded as a hypernym of *appear*.v. This level of generalization would not serve the function of the guidewords. In addition, some verb senses lack hypernyms, e.g., three senses of *appear*.v, and the hypernym-hyponym relation is not applicable to the adjective net. WN entries are the most resistant to modifications given the absence of definitions and guidewords or another feature that can function as a guideword.

5.1.3 Selection of Test Examples

After shortening the entries and presenting them in a familiar dictionary format, the test examples were selected. In this regard, there are at least three possible options for selecting sentences that realize one sense of the target words. The first method is the selection of corpus examples that have one of the senses of the target word. For instance, Nesi and Tan (2011) used test sentences from the British Academic Written English (BAWE) corpus, whereas Dziemianko (2016) used the Corpus of Contemporary American English (COCA). However, this method usually requires manual sense-tagging by at least two lexicographers. This, in turn, needs advanced statistical tests to ensure quality control and consistency among the annotation results. It may also need some editorial decisions to shorten or slightly modify the entry.

Addressing the challenge of manual annotation for a lexicographic study, the researcher may benefit from existing sense-tagged corpora that have already been checked for consistency and quality. SemCor is a promising candidate in this context. However, there are three challenges to using SemCor examples for this test. The annotation is based on WN senses which are more fine-grained than FN's and OALD's senses. Accordingly, several SemCor annotated examples will not find a sense match in the other two resources. There are multiple cases in which the same sentence is assigned two senses because of the uncertainty of the context. In addition, many citations from SemCor are lengthy and contain several unfamiliar words. They are not compatible with the criteria of Good Dictionary Examples (GDEX). Therefore, they may not be beneficial for lexicographic purposes.

Good dictionary examples should reflect frequent usage patterns, be informative and avoid anaphoric references to enhance the readability and the learnability of the example. Therefore, they should be complete sentences, not fragments. The average example length ranges from ten to 25 words. Examples should rely on high-frequency words and avoid using deictic expressions such as *this* and *that* (Kilgarriff et al., 2008).

Therefore, the current study depended on existing dictionary examples as test sentences. This method was also used in the literature (e.g. Ptasznik and Lew, 2014). OALD examples were used to address the challenges of the abovementioned methods and to guarantee that the criteria of GDEX are observed to a great extent. Also, the exclusion of examples from the modified entries facilitated the use of OALD examples as test sentences. The senses recorded in OALD are usually present in the other two language resources. On the contrary, FN and WN provide senses that are absent from other dictionaries. The examples were selected to reflect the sense that is present in the three language resources and that displayed the least degree of overlap with other senses. For instance, *she regularly <u>appears</u> on T.V* was picked to illustrate a common and salient sense of *appear*.v in the three resources (OALD: "to take part in a film, play, television programme"; FN: "to perform in a play or film"; WN: "appear as a character on stage or appear in a play". The target word was replaced with a lost word, as explained in section 3.1. The test sentence that was displayed to the participants was *she regularly <u>avunculizes</u> on T.V.*

5.1.4 Designing Test Items

The study aims at testing the impact of the three entries on the performance of ESL learners. It focuses on five variables relevant to the students' performance, namely, accuracy and time of sense selection, accuracy and time of synonym production and user perplexity. Each variable was tested by a question based on a dictionary entry. The second part of the test asked direct questions about guidewords and definitions, the frequency of using dictionaries, proficiency level and mother tongue[24].

5.1.4.1 Measuring accuracy and time of sense selection

Participants in all groups were exposed to a dictionary entry of one of the five target words. On the same page, there was a sentence including the target word underlined. Then, a multiple-choice question asked learners to choose the correct sense of the underlined word based on the entry they read. They were allowed to choose a single option. This type of question was the most reported method to test this variable in the literature. The answers were scored as follows. If the participant chose the correct sense corresponding to the sense identified by OALD, 1 was assigned to the answer. Otherwise, 0 was given even if users selected a closely related sense.

The software automatically measures the time of sense selection. The software measures the time for each question and for the whole experiment in milliseconds. The reaction time is downloadable in a separate Excel worksheet that is different from the sheet of the recorded responses.

5.1.4.2 Measuring user perplexity

The third variable is user perplexity, which may result from the sense delineation method used in each entry. For instance, OALD lists senses five and six of *development*.n under the same signpost "new buildings". The two definitions overlap (sense 5: a piece of land with new buildings on it and sense 6: the process of using an area of land, especially to make a profit by building on it). Such cases are expected to cause user uncertainty even if the user chooses the correct sense. Therefore, the second question asked in the test was, "Is there any other sense that

[24] The FN-based test is accessible through the following link https://www.psytoolkit.org/c/3.4.4/survey?s=4me5n; the WN-based test is accessible through the following link https://www.psytoolkit.org/c/3.4.4/survey?s=L4PEZ; the OALD-based test is accessible through the following link https://www.psytoolkit.org/c/3.4.4/survey?s=eq2ej.

applies to the use of the underlined word in the example?" and users were allowed to choose more than one option. User perplexity is measured as follows.

- $Perplexity(user_x, word_y) = \frac{number\ of\ selected\ senses}{total\ number\ of\ word\ senses}$
- e.g. $Perplexity(user_1, word_{appear}) = \frac{2}{9}$

5.1.4.3 Measuring Accuracy and Time of Sense Production

The third question was relevant to production skills. Participants were asked to write an English word that had the same meaning as the underlined word in the example. The accuracy of production was measured using the Pedersen et al. (2004) path measure. Accuracy is based on the length of the path between the two concepts in the WN database. The further the two words are in the hierarchy, the lesser the value of the accuracy. It is only applicable to nouns and verbs. If the participant wrote the exact word or a synonym of the word in the target sense, 1 is assigned to the accuracy. If the participant wrote a (direct) hypernym or hyponym, for instance, of the word, 0.5 is assigned to the produced word. For adjectives, synonyms and defining words were assigned 1. Otherwise, 0 was assigned.

5.1.4.4 Perceived Importance of Guidewords and Definitions

The fourth question asked students whether they depended on the definitions or the guidewords in order to identify the correct sense of the word. It aims at evaluating the perceived importance of guidewords and definitions as reported by the participants. Table 5 shows how each option assigns two values for guidewords and definitions.

Table 5. Importance of guide words and definitions

Option	Guide word	Definition
Guide words were enough	3	0
Guide words were more important, but definitions were useful too	2	1
Definitions were more important, but guide words were useful too	1	2
Definitions were enough	0	3

3.1.4.5 Other test items

The second part of the test asked questions to complete the user profile. It included questions about the first language and the proficiency level of the participant. It also asked about their use of dictionaries. One question was asked about the frequency of using monolingual dictionaries, and another was about the frequency of using bilingual dictionaries.

5.2 Test Instrument

Previous lexicographic user-based studies used both paper-based and electronic tests. Paper-based tests usually allocate a space in which the user would subjectively report the time spent on the consultation process (Lew and Pajkowska, 2007), which may not be very accurate. In contrast, electronic tests have the advantage of automatic time measurement. Electronic tests were usually Moodle-based experiments. However, they had to adjust the system to facilitate measuring the time for each test item (Dziemianko, 2016; Nesi and Tan, 2011).

This study used Psytoolkit, which offers more options than Moodle. Psytoolkit is an online toolkit for cognitive and psychological experiments and surveys. It offers precise time measurements for each test item. The web-based toolkit can be used for free to design, launch and analyze questionnaires and experiments. It stores multiple baseline psychological and cognitive experiments. The toolkit can be used to build customized surveys and offer "radio" questions, which are equivalent to multiple-choice questions. It allows participants to select a single choice. For "check" questions, ticking several options is allowed. "Text" questions allow participants to provide a short text in their responses. After data collection, results are downloadable in separate spreadsheets, one for the answers and another for the time spent by each user on each question[25]. Also, an individual plain text file for each participant's response is downloadable (Stoet, 2017).

5.3 Study Hypotheses

The test was designed to explore the following hypotheses.

H_1: The different sense delineation methods in OALD, WN and FN will result in different entries.

H_2: OALD-based, WN-based, and FN-based entries will result in different performances of ESL students.

H_3: The number of senses in an entry correlates with the accuracy of sense selection.

H_4: Entry length correlates with the time of sense selection.

[25] The results of the three experiments conducted in this dissertation are accessible through this link: https://drive.google.com/drive/folders/1S50sUkjUuSEg9wIlfYoPKXXf6W2ExfL4?usp=share_link

H_5: Sense selection correlates with synonym production.

One-way ANOVA and Pearson's correlation tests were used to test the inter- and intragroup differences and correlations.

5.4 Results

The three-task test was answered by 24 participants. According to Cronbach's Alpha, the reliability of the sense selection task was good, the reliability of the perplexity measure was acceptable, and the reliability of the synonym selection task was good. The first task (i.e. sense selection) included five items (α= 0.80), the second task, which measured perplexity, consisted of 5 items (α= 0.70), and the synonym production task also included five items (α= 0.85).

5.4.1 Participants

Twenty-four participants took part in this study. Participants were 1st year English majors at the Institute of English and American Studies, University of Debrecen. Hungarian was the mother tongue of 19 students, while Arabic, Thai, Chinese, Russian and Indian were the mother tongues of five students. The expected proficiency level of the students at this educational stage in English language programs is expected to be at least B2. The self-reported proficiency levels were usually beyond B2. It was evident that participants preferred the use of bilingual dictionaries to monolingual ones, regardless of their first language. The majority of the students reported their daily or weekly use of bilingual dictionaries and their occasional use of monolingual dictionaries. Only a couple of students reported their daily use of both monolingual and bilingual dictionaries. The frequency of using dictionaries did not significantly correlate with either the decoding or encoding performance of students. Table 6 lists the proficiency levels and the frequency of dictionary use reported by the students in the three groups.

Table 6. The proficiency levels and frequency of dictionary use as reported by the participants in Experiment I

		FN-group	WN-group	OALD-group
Proficiency level	Below B2	0	0	0
	B2	13%	25%	13%
	Beyond B2	87%	75%	87%
Bilingual dictionary use	Daily	12%	12%	0
	Weekly	25%	12%	25%
	Monthly	63%	76%	75%
	Bi-yearly	0	0	0
Monolingual dictionary use	Daily	0	12%	0
	Weekly	12%	0	0
	Monthly	25%	25%	12%
	Bi-yearly	25%	12%	50%
	Yearly	38%	51%	38%

Google Translate was dominantly mentioned as the preferred dictionary. Students provided a variety of reasons for their choice, such as "easy to use," "available on my mobile phone," "translate texts," "fast," and "voice translation." Unrestricted availability, quick accessibility, full-text translation, and user-friendliness summarize most of the reasons students gave for their choice. A few students stated some negatively appraised features in traditional dictionaries that were not present in Google Translate, such as long entries in dictionaries. The consultation contexts mentioned by the participants included both academic and non-academic scenes and a variety of encoding and decoding purposes. "Translation" was the main reason for consulting dictionaries. Participants used "translation" as a generic term denoting academic translation, understanding the meaning of English sentences or words through learning their equivalent in their own languages and expressing their ideas in English. The participants did not show any familiarity with FN, WN, or the new types of dictionaries, such as crowdsourced or collaborative dictionaries.

5.4.2 Differences in Word Entries

The study hypothesized that the delineation methods in OALD (intuitive and based on the lexicographer's experience), FN (based on FS and corpus evidence) and WN (based on sense relations) would result in different word entries. The hypothesis was tested through the modified entries of the five previously selected words.

The different sense delineation methods behind the three language resources proved effective in approaching the different entry types. First, the definition length of the target word senses (i.e.,

word count) significantly differs across the three types of entry (F= 32.384, P= 0.00001). For instance, the correct sense of *tell*.v in the three entries obviously varies in length (FN: *instruct someone to do something*; WN: *to request, order, tell, enjoin, say or give instructions to or direct someone to do something with authority*; OALD: *to order or advise someone to do something*).

There were also significant differences in the entry length for the three groups (F= 19.37, P= 0.00001). Entry length is the total word count of the definitions and the guidewords or frame names. The average entry length of FN-based entries is 70 words, whereas it is 94 for OALD-based entries. WN has the longest entries, as the average word count in each WN-based entry was 122 words.

The number of senses identified in each language resource significantly differed, too (F= 7.718, P= 0.001). Despite having a comparable number of senses being one of the inclusion criteria for test words, the differences were statistically significant. For instance, the FN database recognized six to ten senses for each test word, and OALD identified six to nine senses. Similarly, WN recorded seven to ten senses. This dissimilarity would indicate the impact of the sense delineation methods on the number of the realized senses in each language resource, even if the differences were not apparent from a descriptive perspective. Accordingly, the results of the one-way ANOVA test support the first hypothesis of the study.

Each of these variables (i.e., entry length, definition length and the number of senses) correlated differently with the student-dependent variables (e.g., time of sense selection, correct sense selection and user perplexity) across the three tested groups.

5.4.3 Accuracy of Sense Selection

The second hypothesis postulated in the study assumed that the different entries examined in the test would result in different performances of ESL students. This includes accuracy and speed of sense selection, user perplexity and accuracy and speed of synonym production.

Accuracy of sense selection was measured through the total number of correct sense choices among participants in each group. The three groups provided significantly different responses when it came to the accuracy of sense selection (F= 4.089, P= 0.0211). The result is significant at P< 0.05. The result supports the second hypothesis of the study.

As visualized in Figure 8, students who examined FN entries were the most successful in choosing the correct senses (58%). *Autexousious*.n, the substitute for *fair*.n, was the most

successfully consulted entry. All students in the FN group were able to select the correct sense of *fair*.n in the test sentence. The CEFR level of *fair*.n is A2. The correct senses of *avunculize*.v and *aeipathy*.n, which replaced *appear*.v and *level*.n, were identified by 62.5% of the students, although their CEFR levels were A1 and A2, respectively. Only 37% and 25% of the participants in the FN group could correctly identify the sense of *development*.n (B2) and *tell*.n (A1) respectively.

Students who consulted WN entries showed a relatively lower performance (55%) than those who saw FN entries. The accuracy of sense selection for *level, development* and *fair* dropped in the WN group as compared to the FN group but was the same for *appear*.n. Interestingly, the correct sense of *tell*.v which was the least successfully identified by the FN group (25%) was correctly identified by 75% of the WN group. It was the single entry for which the WN group displayed considerably better performance than the FN group.

Although OALD provided baseline entries in lexicography, the group that consulted OALD-based entries showed the lowest decoding performance (46%). The accuracy of sense selection saliently decreased for each test word. Even the sense of *fair*.a, which was the most successfully selected by WN and FN participants, was correctly identified by only 50% of the participants in the OALD group.

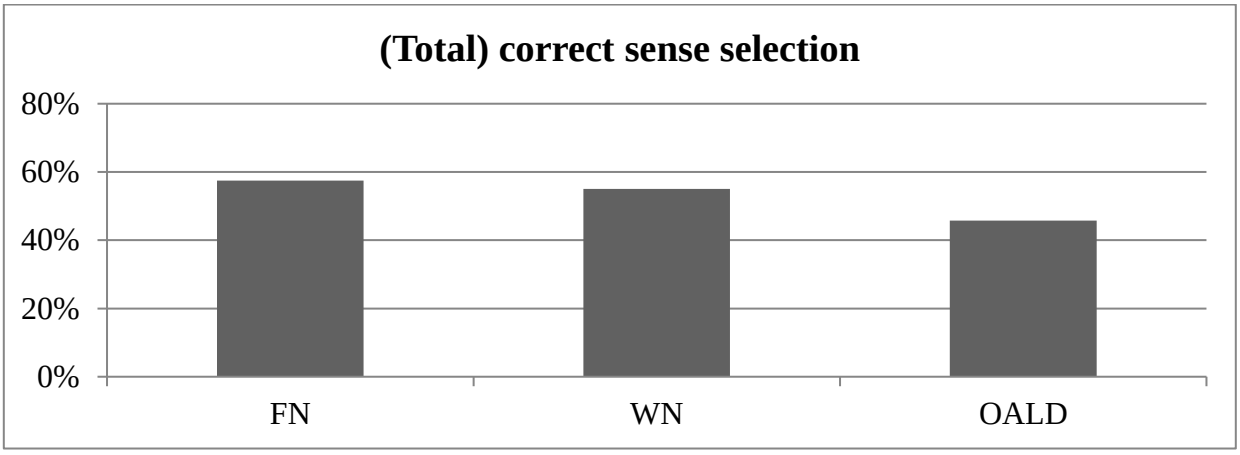

Figure 8. Accuracy of sense selection among FN, WN and OALD groups

The third hypothesis (i.e., the number of senses correlates with the accuracy of sense selection) was not supported for the three groups. On the contrary, another entry-related variable (i.e., definition length) showed significant negative correlations with the accuracy of sense selection

for the WN and FN groups. For the WN group, the length of the definition of the correct sense negatively correlated with correct sense selection (r= -0.452, P= 0.00341). The significant negative correlation indicates that when the definition was longer, students were less successful in the identification of the correct sense. Inaccuracy may be caused by information overload in the definition. For the OALD group, the correlation between the definition length and the accuracy of sense selection was positive but not statistically significant. Definition length and correct sense selection revealed a significant negative correlation for only the FN group (r= -0.362, P= 0.0217). The correlation was negative but not significant for the WN and the OALD groups.

5.4.4 Time of Sense Selection

Figure 9 shows the sense selection time for the three groups. The time of sense selection also significantly differed across the three groups (F= 3.58, P= 0.033). This significant difference, again, supports the second hypothesis of the study.

Participants in the FN group spent the shortest time consulting the entry and selecting the sense. On average, they spent more than one minute on each word. There were salient differences in the time spent on the different and same word entries by the multiple FN participants. For instance, the shortest time recorded was 24 seconds for the *level*.n entry in the FN group, whereas the longest time (two minutes and 36 seconds) was recorded for consulting the entry of *appear*.v.

Participants in the OALD group spent more time on the consultation process. They took more than 70 seconds on average to read the entry and choose the correct sense. The longest response time was three minutes and 56 seconds, and it was recorded for the word *appear*.v. The shortest time recorded was 12 seconds for *tell*.v. Worthy mentioning, *tell*.v had eight senses grouped under seven guidewords in the OALD entry. Spending 12 seconds reading the entire entry, the question and the test sentence and selecting the answer does not seem a plausible possibility. This strongly suggests that the participant skipped several parts of the entry.

Students who read the WN-based entries spent the longest time among the three groups. They took on average 77.9 seconds to read the entry and choose the correct sense. The shortest time was 30 seconds, and it was associated with *level*.n. The longest time was three minutes and 16 seconds for the word *appear*.v.

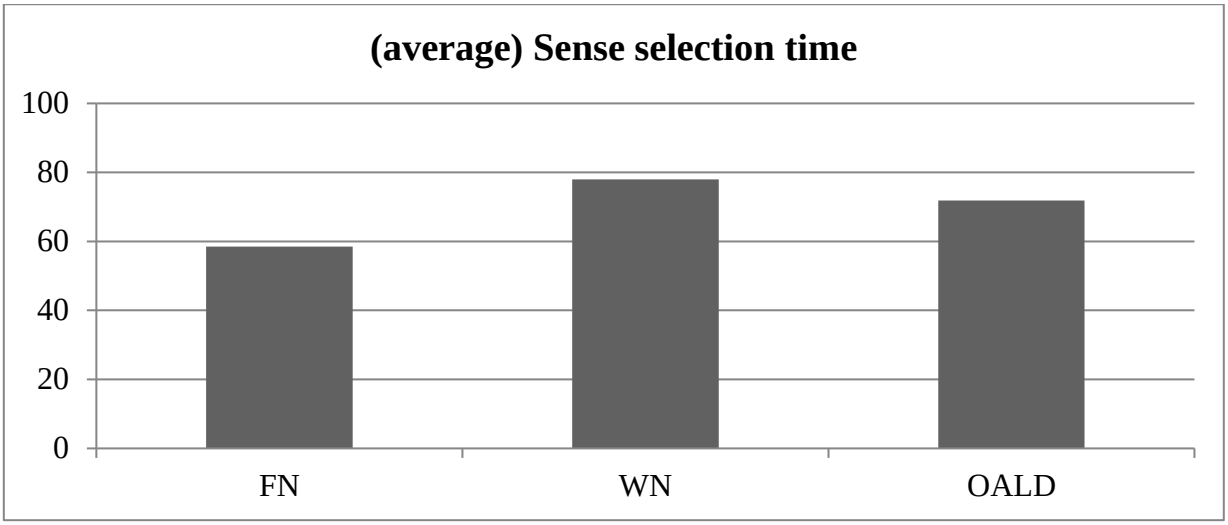

Figure 9. Time of sense selection among FN, WN and OALD groups

For the FN group, three variables were significantly correlated with the time of sense selection. First, the entry length, which is the total number of words in the entry (including definitions and guidewords), was positively correlated with the time of sense selection (r= 0.449, P= 0.0036). Second, there was a positive correlation between the number of senses and the time of sense selection (r= 0.423, P= 0.007). Finally, the length of the definition and the sense selection time were positively correlated (r= 0.393, P= 0.012). Accordingly, when participants encountered more information in the entry or in the definition of the target sense, they spent more time on the consultation process.

For the OALD group, definition length, number of senses and entry length showed a positive correlation with the time of sense selection. However, the results were not statistically significant.

On the contrary, the number of senses, definition length and entry length in WN-based entries negatively correlated with the time of sense selection. However, the anti-correlation was significant only for the number of senses (r= -0.484, P= 0.0016). That is to say, the longer entries students had, the less time they spent on the consultation process. Although the correlations between the same variables (i.e., number of senses and time of sense selection) were measured in FN and WN groups, and both were significant. The first was positive, whereas the second was negative. This may suggest that the WN group skipped parts of the entry, but the FN group did not. Hence, the fourth hypothesis was verified only for the FN group.

5.4.5 User Perplexity

Participants in the three groups also revealed significantly different levels of user perplexity (F= 23.278, *P*= 0.00001). Perplexity levels were measured for cases of correct sense selection only. That is to say, the perplexity that was associated with the inability to identify the correct sense was discarded. The study was more concerned with competent participants who managed to choose the correct sense, but they still thought other senses could be applicable to the sense used in the test sentence.

FN-based entries caused the slightest degree of user perplexity. Usually, participants were perplexed by one sense of the target word or not perplexed at all. The only exception was the entry of *appear*.v. All of the participants who could correctly identify the right sense of the word were confused by at least one or two senses that were different from the sense they chose. *Fair*.a and *level*.n caused the least degree of perplexity for users. Only 25% of the users thought that another sense could be applicable to the use of *fair*.a in the test, whereas 20% of the participants believed that two other senses were appropriate for the use of *level*.n in the test sentence.

WN-based entries were associated with higher levels of perplexity. On average, participants thought the target word corresponded to at least two senses in the provided entry. For *appear*.v, *tell*.v and *development*.n, all participants were perplexed by at least one sense and a maximum of four senses in some cases. Similar to the FN group, *level*.n and *fair*.a resulted in the lowest perplexity levels (33% chose one more sense for *level*.n whereas 50% checked one or two senses for *fair*.a).

The group that was exposed to OALD entries revealed the highest perplexity levels. Of the participants who selected correct responses, 56% also selected other senses as potential meanings of the target word. *Appear*.v and *development*.n were the most perplexing to users. All participants who chose the correct sense of the two words checked at least one extra sense as a possible candidate for the meaning evoked by the test sentence. Moreover, 50% of the participants were perplexed by the entry of *level*.n, which was the most successfully consulted entry for this group.

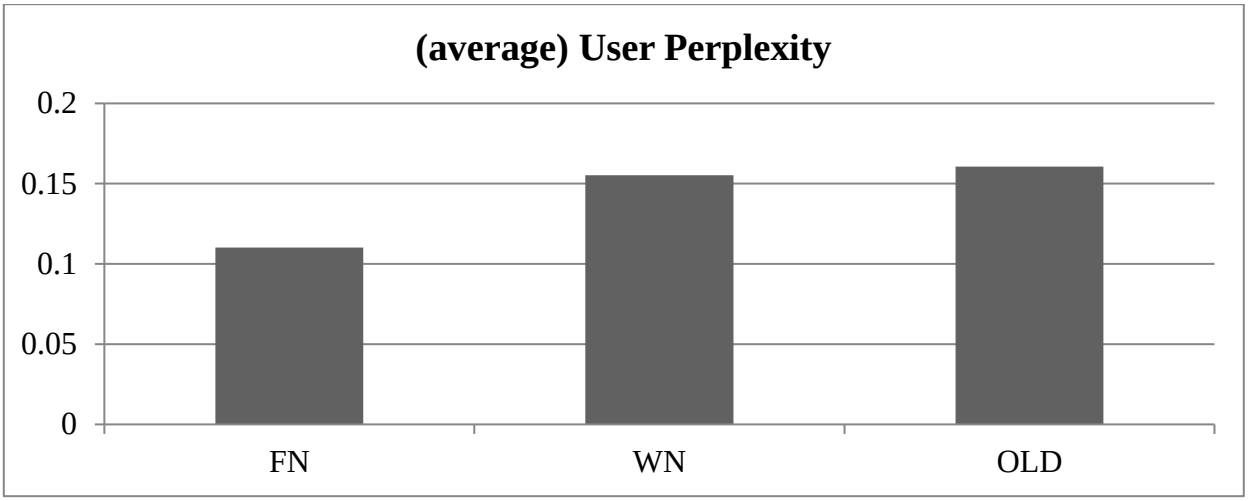

Figure 10. User perplexity scales among FN, WN and OALD groups

For the three groups, the was a negative correlation between the accuracy of sense selection and user perplexity, but the results were not statistically significant. Also, there was a positive correlation between the time of sense selection and user perplexity among the three groups, but the results were not statistically significant. There was a positive correlation between the number of senses and perplexity in the FN and OALD groups, but the results were not statistically significant too.

However, the correlation between the number of senses and perplexity levels was negative in the WN group (r= -0.471), and it was the only statistically significant correlation (P= 0.002) between perplexity and another variable. Participants were less perplexed when they read a larger number of senses. Although this seems surprising, this is compatible with the previously reported negative correlations between the number of senses and the time of sense selection. This, again, may suggest that participants in the WN group did not examine the entire entry and, accordingly, were not very perplexed by the relatively increasing quantity of information in each entry.

5.4.6 Accuracy of Synonym Production

The last task students were asked to perform was a synonym production task. There were no statistically significant differences among the participants in all groups. The FN group produced the most accurate words that were similar to the target word. At least 25% of the participants were able to write the original target word for each test word. Synonyms of the target sense also frequently appeared, e.g., *pale* and *light-colored* for *fair* (37.5%) and *floor* for *level* (25%). Also,

hyponyms of some words were present in the student's production, such as *perform*.v for *appear*.v (37.5%). In some cases, students wrote a troponym of the verb, such as *instruct* for *tell* (12.5%).

The second-best production performance was given by the WN group. The presence of paradigmatic sense relations was more evident among the participants in this group than in other groups. The synonym *floor* accounted for 62.5% of the production for *level*.n and 50% of the production for *fair*.a was *light-colored.* Also, direct and indirect hypernyms were salient in the produced words, e.g., *warn* was 50% present as a substitute for *tell*.v. The poorest performance was relevant to the word *development*.n. Only 25% of the participants managed to unveil the original word in their production, whereas other responses were relevant to other senses of *development*.n.

Participants in the OALD group were the least successful in writing a word similar to the original target word. The most relevant production was for the word *fair*.a as 50% of the students managed to write *pale* as a word having the same meaning as *fair*.a. However, other words were irrelevant (e.g., 25% wrote *amazing* or *wonderful*). In the same vein, 25% could write the original word *appear*.v for the first production task, however, other words were barely relevant in terms of sense relations (*show off*, *broadcast* and *show* constituted 50% of the production).

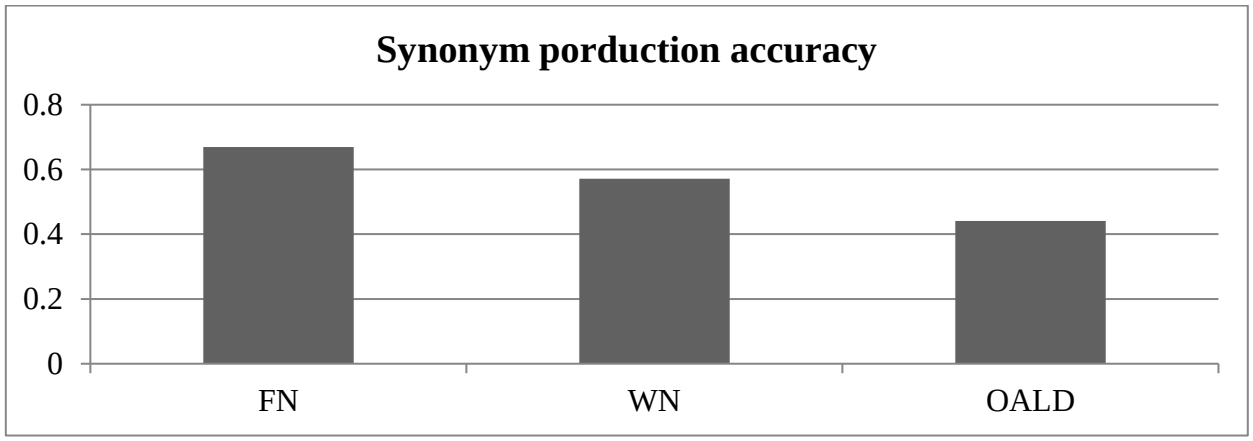

Figure 11. Accuracy of synonym production

There was a positive correlation between correct sense selection and correct synonym production in the four groups. The correlation was only significant in the WN group (r= 0.354, P= 0.037). That is to say, the fifth hypothesis was supported for the WN group only.

5.4.7 Time of Synonym Production

Time of synonym production also varied across all groups. The FN group spent the shortest time producing a word that had the same meaning as the original word. Participants spent less than 17 seconds on average to think about a word and write it. In this question, they only had to read the sentence they had already read and chosen the correct sense for. This explains the very short time spent on this task and, at the same time, the successful production. The shortest time for producing a 100% correct answer was 2.8 seconds, and the participant could write the exact original word *level*.n. The longest time was 106.019 seconds for a relatively correct production *to appear somewhere.*

The OALD group spent a relatively longer time than the FN group on the production task. They took an average of 20 seconds to write a word with the same meaning as the target word. However, most of the production was unsuccessful. The shortest time recorded was also for the production of a similar word to *level*.n. The participant took 4.786 seconds to produce the synonym *floor*. The longest time a participant spent to produce a similar word to the target was 132.5 seconds to write *act* as a similar word to *appear*.v (25% similarity between the two words).

Like the time of sense selection, the WN group took the longest time to produce a synonym of the target word. Students spent 27.7 seconds on each production task on average. The least time spent to find an equivalent word was 4.8 seconds, and the attempt was successful (i.e., *floor* for *level*.n). The longest time was 86.327 seconds, and it was recorded for the word *play* as a substitute for *appear*.v, but the production was less successful (*play* is 25% similar to the sense of *appear*.v in the test sentence).

That is to say, the shortest time was relevant to the word they saw at last, while the production of a word that was similar to the first word on the test took the longest time for the three groups. This may suggest increasing familiarity with the task and expectancy of the question based on previous exposure (Figure 12).

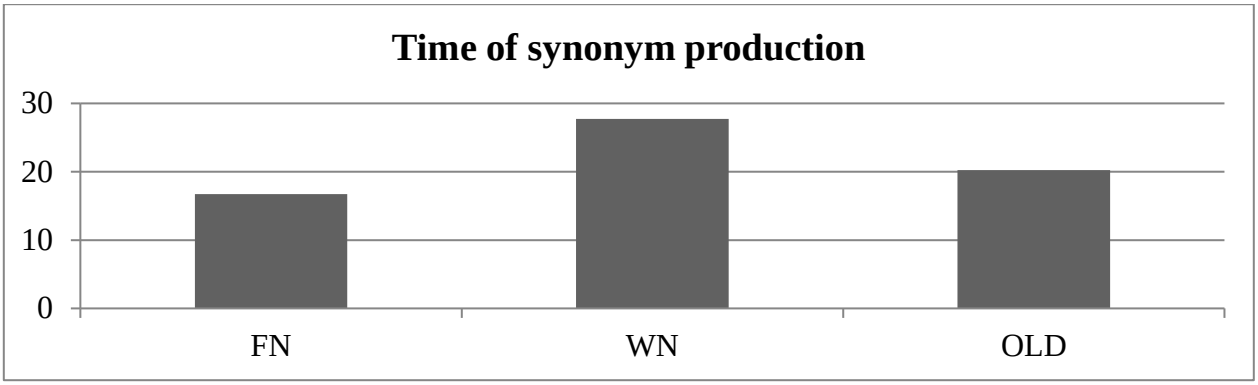

Figure 12. Time of synonym production

For the three groups, there was a negative correlation between the accuracy of synonym production and the time spent on producing the synonym. The results were not statistically significant, though. There was also a negative correlation between the time of synonym production and sense selection for the OALD group, but the correlation was positive for the other two groups. The results were altogether statistically insignificant.

5.4.8 The Perceived Importance of Guidewords and Definitions

Guidewords were available only for FN-based and OALD-modified entries. For FN-based entries, frame names were generally more complex than traditional OALD guidewords because they refer to a situation that can be evoked by multiple LUs. On the contrary, OALD guidewords were usually a synonymy, hypernym or typical subject or object of the lexical unit. Therefore, considerable discrepancies can be detected between the two types of information used as guidewords. For instance, "floor or layer" is the OALD's guideword for the second sense of *level*.n (i.e., *a floor of a building; a layer of ground*). The equivalent sense in FN is included in the BUILDING_SUBPARTS frame. Also, OALD's guideword for the last sense of *fair*.a is "beautiful", which is both the definition and the guideword. FN's equivalent frame is AESTHETICS.

In this regard, participants in the OALD group reported their reliance on the guidewords more than on the definitions to identify the correct sense of the word. Figure 13 sums the scores for the importance of guidewords and definitions according to the participants in the two groups. Unlike the OALD group, participants in the FN group reported their dependence on the definitions more than the guidewords to choose the correct word sense. This may indicate the students' extra effort to process the names of frames compared to their counterparts who read the OALD entries.

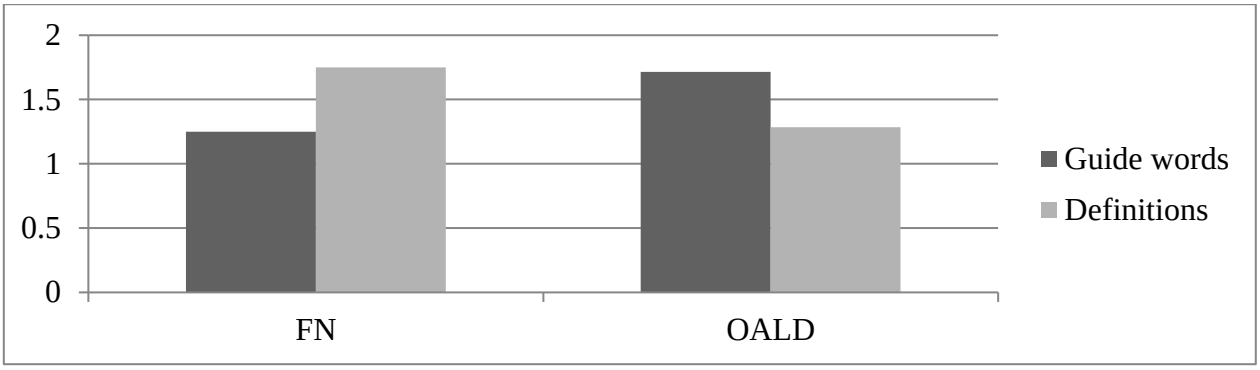

Figure 13. The perceived importance of guidewords and definitions in FN and OALD groups

5.5 Discussion

Finding the right sense in a long polysemous entry has always been challenging for learners. Learners frequently stop at the first couple of senses in long entries and terminate the consultation process. Therefore, various studies have explored the influence of the number of senses, entry length and definition length on the accuracy and speed of sense selection.

The influence of the number of senses in a dictionary entry in the present study was significantly and positively correlated with the time of sense selection for the FN group. This is compatible with Lew and Pajkowska's (2007) finding that longer dictionary entries (which displayed about seven senses) slowed down the consultation process by 60% if compared to shorter entries. Also, in their translation task, the accuracy improved by 22% for short entries. However, similar to the present study, the number of senses did not correlate with the accuracy of sense selection.

The results of the present experiment displayed a positive and statistically significant correlation between definition length, entry length and time of sense selection among the participants in the FN group. This is partly compatible with the results of Nesi and Tan (2011), which found significant positive correlations between the length of the definition and time of sense selection but did not find a correlation between entry length and time of sense selection. Also, Ptasznik and Lew (2014) confirmed that participants spent longer consultation time on entries that had more senses. To elaborate, students spent approximately 38 seconds on entries with five senses, 44 seconds on entries with seven senses and 49 seconds on entries that had nine senses.

Unlike before, Ptasznik and Lew (2019) could not detect any significant correlation between entry length and time of sense selection. On the contrary, they found "a hint of a reverse trend".

That is to say, students tend to spend less time on longer entries. This surprising "hint" is somehow consistent with the results of the present experiment regarding the anti-correlation between entry length and time of sense selection in the WN group.

Level.n was the most successfully identified word, in the shortest time and with the lowest degree of perplexity. All the senses share the same conceptual base in which space and verticality are essential domains. Whereas some senses evoked measure, quantity and size (e.g., amount of something), others were relevant to hierarchy and evaluation (e.g., position or importance). The tested sense was more concrete as it denoted a vertical level with reference to a building. The (metaphoric) extensions of most of the senses from the basic sense (i.e., points on a vertical space) can be easily detected. The detectable links between the different senses and the concreteness of the tested sense may have been factors in the high performance associated with *level* in the three groups.

Appear.v was associated with the least successful answers, longest consultation time and highest perplexity across the three groups. Examining the entries of *appear* in the three groups shows how fuzzy categorization is applicable to its senses. *Appear* has a very complex meaning as it involves knowledge about visual experience. Visual experience is an archetype which is very basic but very complex to explain. It typically includes a perceiver, an entity or a phenomenon and perception through the visual modality. Although this conceptual base is shared with verbs of *seeing*, *appear* is different in its focus on the phenomenon that became visible, not on the perceptual experience or the perceiver. This conceptual base can explain the metaphoric extension of the meaning of *appear* as *look, seem* or *give an impression*. This sense focuses on an entity or a phenomenon (which is not necessarily perceived by the senses) but is evaluated or appraised by an implicit or explicit perceiver or cognizer. This is consistent with the conceptual metaphor ABILITY TO SEE IS ABILITY TO EVALUATE. This metaphoric extension may be the easiest to delineate and link to the basic sense of *appear* as "be seen". Other senses are more challenging to categorize. For instance, "begin to exist" seems to be a more basic sense than "to be seen" because seeing something entails its existence but the existence of something does not necessitate its perception by visual modality. The link between these two senses does not correspond to Langacker's cases of elaboration or extension or Fillmore's shared FEs or frame relations. There have been cases in which the meaning of a word is centered upon more than one basic sense. However, for *appear* accepting "be seen" and "begin to exist" as two basic senses

would exacerbate the challenge of categorizing the rest of the senses. The sense of *appear* as "publishing a book or an article" is related to both the *existence* and *seeing* senses. It would not be plausible to delineate it as an extension of one of the senses and exclude the other. The same is applicable to the sense tested in the experiment. To *appear* on TV is to perform or play a role. This is relevant to both the metaphoric existence in media and the meant-to-be-seen performance.

The influence of target sense position was also examined in the literature. The salience of the first senses and the discard of the last senses in the consultation process are reiterated in many studies. However, the results of some empirical studies opposed such arguments. For instance, Nesi and Tan (2011) reported surprising results regarding the relation between sense position and speed and accuracy of sense selection. They stated that "the last sense in the entry proved easiest to identify". Participants in their study were able to identify the last sense with greater accuracy and speed than the middle senses. Also, Dziemianko (2016) reported significantly faster identification of final position senses (85 seconds) if compared to initial position senses (97 seconds). However, the influence of sense position on accuracy selection, meaning reception or production was not detected (Dziemianko, 2017).

The influence of signposts on the speed and accuracy of dictionary consultation has been widely discussed in the literature. Signposts, guidewords or shortcuts are fundamentally introduced to help learners navigate longer entries and identify the correct word sense in a shorter time. The present study did not directly examine the influence of guidewords on the performance of students. It did, however, explore their perceived importance, as reported by students. There was no significant correlation between the perceived importance of guidewords and the performance of students. It is worth mentioning that WN-based entries, which had no signposts, were associated with the longest consultation time.

This indirectly supports the previous arguments for the role of signposts in shortening consultation time. In this regard, Lew and Pajkowska (2007) proved that the presence of signposts speeds up the consultation process. In their study, the group which consulted dictionary entries with signposts was 14.4% quicker than the control group. However, signposts did not display a significant correlation with the accuracy of sense selection or the accuracy of the translation. Similarly, Ptasznik and Lew (2014) mentioned that the presence of signposts and signposts and menus reduced the time of sense selection from 49 seconds in the control group

(entries without signposts) to 40 and 42 seconds, respectively. Signposts also slightly improved the accuracy of sense selection, but the result was not statistically significant.

On the contrary, Nesi and Tan (2011) reported other results regarding signposts and time of sense selection. Their experiment concluded that there were no significant differences in the consultation time between all groups who consulted entries with signposts, entries with menus and entries without signposts. However, there was a significant improvement in the accuracy of sense selection for the entries that had signposts compared to those without signposts.

In addition to the presence and absence of signposts, some studies were interested in the influence of the visualization and phrasing of the signposts. Dziemianko (2016) identified a statistically significant difference in the influence of the visualization of signposts on the time of sense selection. Specifically, white capitals on a blue background were the visualization that resulted in the shortest consultation time (i.e., 81 seconds). Dziemianko (2019) referred to the significant effectiveness of heterogeneous signposts (i.e., belonging to different grammatical classes) in sense selection and reception if compared to homogeneous signposts. However, the use of homogeneous signposts was more effective in enhancing the production performance of students.

At the cognitive level, the signposts used in the FN entries refer to the entire schematic situation associated with each sense. Therefore, they are expected to show the general similarities and, or differences in the scenes evoked by the senses. However, the complexity of phrasing the name of frames impeded their effective use by the participants. The FN team does not employ crowdsourcing techniques to maintain the precision and integrity of the database at the cost of coverage and update. However, crowdsourcing alternative frame names may improve the usability of the database without risking the precision of the lexicographic information.

5.6 Conclusion

To conclude, this experiment aimed to answer two questions. First, it inquired whether the different sense delineation methods in OALD, FN and WN would influence the encoding and decoding performances of ESL learners or not. The statistically significant intergroup differences in the performance confirm such influence. The influence of the different methods was salient in the variations in sense selection accuracy, consultation time and user perplexity. However, there was no statistically significant difference in the production task.

Second, the experiment aimed at identifying the optimal approach for improving the encoding and decoding performances of ESL learners. In this regard, the FN approach retrieved the best results at all levels. Students in the FN group were more accurate in identifying the correct sense, faster in the consultation, less perplexed and more accurate in producing a synonym than the other two groups.

The results have several implications. Although several studies were concerned with measuring the speed and accuracy of sense selection, none of them attempted to compare the influence of different traditional lexicographic resources (e.g., OALD) and theoretically-motivated databases (e.g., FN and WN) in improving the performance of learners. In addition, they were interested in salient dictionary features such as signposts, not in the results of meeting a lexicographic challenge (i.e., sense delineation). The results of this experiment proved that theory-based approaches to sense delineation in WN and FN are more effective in helping ESL students perform decoding and production tasks. This finding supports the argument for the role of linguistic theory in lexicography (Atkins and Rundell, 2008).

In addition, the FN-based approach is proved to be the most successful regarding the accuracy and speed of sense selection and reducing user perplexity. This is compatible with the studies that recommend cognitive semantics as a theoretical framework for lexicography (Geeraerts, 2010; Ostermann, 2015) and as a tool for enhancing the understanding and presentation of senses (Halas, 2016). What distinguishes the FS approach from other cognitive approaches (e.g., PT and CMT) is its relatively more explicit criteria of sense separation, reliance on corpus data and current implementation in a lexical database. This facilitates using it in both lexicographic and teaching contexts.

Chapter Six – Proposing a Hybrid Entry based on FN, WN and OALD

6.1 Methods and Data Collection

This chapter draws on the results of Experiment I. It proposes a new hybrid entry based on the positively evaluated features (statistically and qualitatively) inferred from using FN, WN and OALD entries. Experiment II aims at constructing a theoretically-motivated dictionary entry that elicits successful decoding and encoding responses from ESL learners. Therefore, this experiment addresses the following questions:

1. What are the challenges of integrating linguistic information from OALD, FN and WN in a single hybrid entry?

2. Does the proposed hybrid entry elicit better decoding and encoding performances from ESL learners than the traditional OALD entry?

Experiment II evaluates the same variables examined in Experiment I, uses the same type of questions and assigns the same tool. The only test item that slightly differed in Experiment II was relevant to the perceived importance of guidewords. A fifth option was included to check if guidewords and definitions were perceived as equally important to sense selection. Experiment II expands the sample of words to include 12 words, and it targets a larger population of ESL learners. The following sections illustrate the new entry design, explain the expansion of the tested words, list the hypotheses and display the results.

6.1.1 Selection of Test Words and Sentences

Seven words were added to the previously tested five words. Experiment II tests twelve words with three parts of speech. The same method of disguising the target words was applied to this experiment[26]. Table 7 displays the twelve words, their replacements and the year in which the replacement word appeared.

[26] The test answered by the target group is accessible through the following link https://www.psytoolkit.org/c/3.4.4/survey?s=n8EdG; the test answered by the control group is accessible through this link https://www.psytoolkit.org/c/3.4.4/survey?s=6nOvj

Table 7. Target words, replacements and years of use

Target word	Replacement	Years of use
Appear.v	Avunculize.v	1662
Blow.v	Saburrate.v	1623 -1658
Break.v	Temerate.v	1635 -1654
Case.n	Scaevity.n	1623 -1658
Cool.a	Noscible.a	1654 -1654
Development.n	Assectation.n	1656
Fair.a	Autexousious.a	1678
Full.a	Ovablastic.a	1922
Level.n	Aeipathy.n	1847 to 1853
Sound.n	Nidifice.n	1656
Strong.a	Senticous.a	1657
Tell.v	Addecimate.v	1612 -1755

A strict application of the criteria for selecting words (mentioned in Experiment I) was relatively challenging for this experiment. The realization of the differences in the sense delineation methods was more salient with the expansion of the sample words than in Experiment I. In addition, the incomplete database of FN imposed further challenges.

First, the previous challenge of having at least five senses for each word was more problematic during the sample expansion. Several words were excluded because they possessed less than five senses in FN. It is worth mentioning that the FN database has 2407, 5575 and 5213 word senses for 2276 adjectives, 4647 nouns and 3274 verbs, respectively. That is to say, the average number of senses for each word ranges from one to two senses. Therefore, the selection of twelve words having at least five senses from FN is a challenging task, primarily if they should be equally distributed across the three word classes defined above.

Second, the criterion of having a comparable number of senses in the two language resources imposed further challenges. In several cases, the number of senses in FN and OALD was considerably distinct. *Cover*.v, *power*.n and *charge*.n, for example, have five recorded senses in the FN database. However, they have fifteen, sixteen and twelve senses in OALD. This diversity might affect the time of consultation taken by each group. Therefore, these cases were excluded from the experiment, along with thirty similar cases.

Third, the criterion of sharing the maximum number of senses between the two resources was waived to having at least two shared senses. The manifestations of the different sense delineation methods and the dissimilar approaches to splitting and lumping senses were also more noticeable in this experiment than in Experiment I. Several cases have a relatively similar number of senses in FN and OALD, but they also were excluded for qualitative reasons. *Find*.v, for instance, has

nine entries in FN and 11 entries in OALD. However, the way the senses are realized and separated in the two dictionaries was too different to find two sentences that can be mapped to the same sense in the two resources. On the one hand, the first four senses in OALD denote the way by which something was found (i.e., by chance, by searching, by studying/searching and by experiencing/testing as reflected in the guidewords). On the other hand, parallel senses in FN instantiate the finder (i.e., cognizer or perceiver) and the thing that is found (i.e., idea, evaluee, goal or location). *Find*.v in BECOMING_AWARE is defined as "to perceive a phenomenon, either by chance or deliberately". This definition neither corresponds to the OALD definitions nor directly lumps the first four senses altogether because FN defines another sense of *find*.v as "[a] cognizer arrives at an idea" without specifying how it was found. One-to-one and one-to-many mappings between the senses recorded in FN and OALD are not possible because they go beyond splitting and lumping techniques.

Accordingly, the selection of test sentences was likewise problematic. The chosen sentence should clearly express a sense that is salient in OALD and FN. Apparently, it was never the first sense. The sentence must realize a sense that expresses a relatively less common meaning than the general sense of the word in order to guarantee that it will be separated from other senses. The senses of *case*.n were similar in the two projects, but OALD provided finer distinctions in the sense of *case* as a "situation". Despite the similarity, several sentences were excluded because the definitions of the senses were not mutually exclusive.

The sixth sense of *case*.v in OALD was "a set of facts or arguments that support one side in a trial, a discussion". *The case against her was very weak* is provided as an instance of this sense. If used as a test sentence, the fourth and the fifth senses in OALD, which refer to a case in "police investigation" and "in court" would apply, too. Therefore, indecisive examples were discarded to minimize user perplexity. For example, the sense of *case*.n as a container is not expected to be confused with other senses. Accordingly, its example in OALD (i.e., *The museum was full of stuffed animals in glass cases*) was selected as the best test example for Experiment II. Similar decisions were taken to select other example sentences for the newly used seven entries. Table 8 illustrates the test sentences for each sense.

Table 8. Target words, senses and test sentences for Experiment II

Target word	Target sense in FN	Test sentence
Appear.v	to perform in a play or film	She regularly avunculizes on TV
Blow.v	completely fail to achieve	She saburrated her chances by arriving late for the interview
Break.v	fail to observe (a law, regulation, or agreement)	He was temeratingt he speed limit
Case.n	a container designed to hold or protect something	The museum was full of stuffed animals in glass scaevity
Cool.a	calm and collected	Just stay noscible and don't panic
Development.n	a large group of buildings constructed to form a community	He bought the land for assectation
Fair.a	pleasurable to the aesthetic senses	Kate had dark hair and autexousious skin
Full.a	containing many objects, but not necessarily without room for more	The sky was ovablastic of brightly coloured fireworks
Level.n	a story or vertical subpart of a building	Take the elevator to aeipathy four
Sound.n	a style of a piece of music	
Strong.a	exceptionally capable	The play has a very senticous cast
Tell.v	instruct (someone) to do something	There was a sign addecimating motorists to slow down

6.1.2 The Proposed Hybrid Entry

The results of Experiment I showed that the FN-based entries helped students select the correct senses more than the two other systems did. The only shortcoming of the FN-based entries was relevant to the importance of guidewords (i.e., frame names in that case). Students reported that definitions were more beneficial than guidewords in the FN group, whereas guidewords were perceived as more valuable by the OALD group. The WN group showed the second-best performance after the FN group. The influence of the additional hypernym and hyponym information in WN entries was reflected in the production task. Moreover, the WN group was the only group for which a significant positive correlation between the accuracy of sense selection and synonym production was observed.

Based on the previous results, the modified entries for this experiment maintained the FN system of identifying and separating senses, replaced frame names with OALD guidewords and added WN's hypernyms and hyponyms based on Semcor frequencies. Figure 14 shows the hybrid entry for the word *appear*.v. The entry contains all FN senses for the target word in addition to their FN-recorded definitions. The OALD guideword precedes each sense. The WN-based information is added in smaller italic font as an extra piece of information that is not part of the definition. The generic term "related words" is used to avoid the precise terms that presuppose the user's familiarity with a particular sense relation (e.g., hypernym, hyponym, troponym).

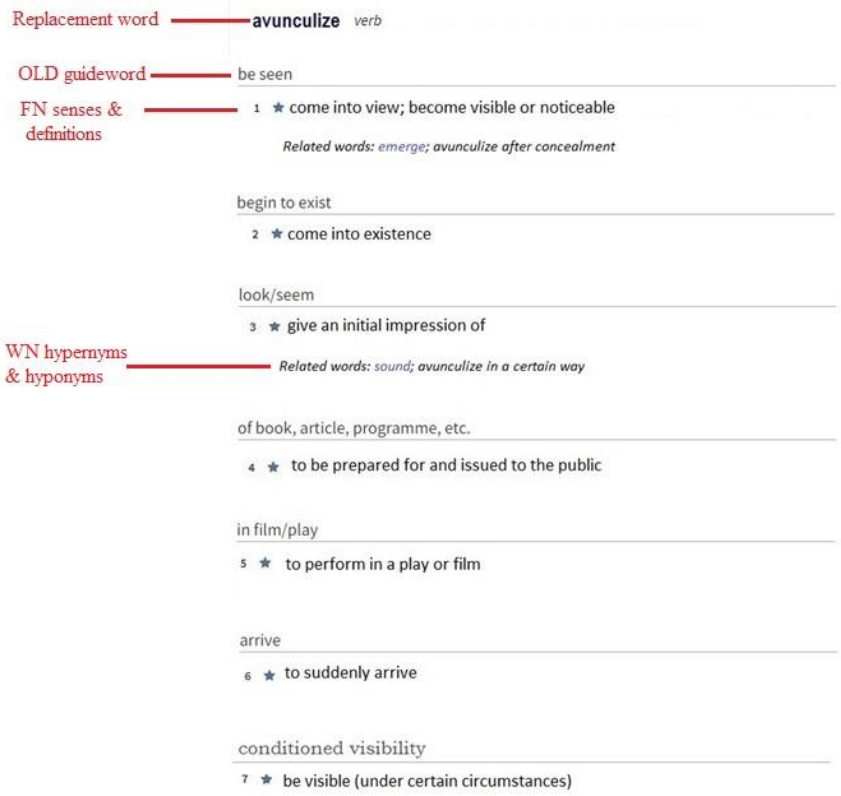

Figure 14. Hybrid entry based on FN, OALD and WN

The choice and phrasing of guidewords in OALD are simpler and more straightforward than those of frame names in FN. However, as discussed in the previous section, there is neither one-to-one nor one-to-many correspondence between FN and OALD senses. Therefore, replacing frame names with guidewords was challenging. The following problems appeared when assigning OALD guidewords to FN senses. First, some FN senses were missing from OALD. In such cases, the original frame name was kept if it was simple and representative of the senses. A simplified version of the frame name was used for clarity if the frame name was representative of the sense. A new word that imitates OALD's guideword was proposed if the frame name was neither representative nor straightforward.

To elaborate, FN realizes a sense of *full*.a as "experiencing emotion", which involves an experiencer *full of* emotions (e.g. *she was full of joy*). This sense was missing in OALD. It is evident that the only signal of the abovementioned FN sense is one of the multiple examples provided for the second sense of *full*.a in OALD. OALD lumps several senses of *full*.a in this

definition “having or containing a large number or amount of something/somebody”. It does not consider the container or the contained. Heterogeneous examples, including *the sky was full of brightly coloured fireworks*, *he is always full of energy* and *life is full of coincidences*, represent this sense in OALD. These senses are separated in FN and, accordingly, the OALD guideword “having a lot” would not match FN senses. In this case, FEELING, the frame evoked by the missing sense of *full*.a (i.e., experiencing emotion), is used as a guideword for its simplicity and representativeness. Other occasions required rendering a guideword that was similar to OALD guidewords because the frame name was too general to help users find the relevant sense. In the PART_WHOLE frame, FN defines *development*.n as “the section of a piece of music where the major musical themes come together”. OALD mentions a single potential sense which could hardly be similar to the FN sense (i.e., the third sense denoting the process of producing or creating a product or an idea). “New product or idea” is the OALD guideword for this sense, but it would be odd to refer to a piece of music as “product” or “idea”. Moreover, a similar sense to that of OALD is already recorded in FN under the PRODUCT_DEVELOPMENT frame. In this case and in similar cases, a guideword (e.g. *music*) was introduced to reflect the specificity of the sense and differentiate it from other senses.

Second, some FN senses were grouped under the same guideword in OALD if it was both simple and representative. However, if it was adopted in the modified entry, the main criterion of separating senses in FN would be negatively manipulated. Whenever FN records two senses of the same word (in the same word class), two frames have to be realized. Therefore, the modified entries could not contain two senses under the same guideword. OALD places both “something that you can hear” and “continuous rapid movements (called vibrations) that travel through air…” under the same guideword “something you hear” in the entry of *sound*.n. In contrast, FN separates the equivalent senses in SOUNDS and SENSATION frames, respectively. Accordingly, the OALD guideword replaced the frame name for the first sense, and a new word (i.e., vibrations) is used for the second sense of *sound*.n. Despite the challenging cases, OALD guidewords were generally beneficial and promising substitutes for frame names. Table 9 presents some of the decisions taken during the substitution of FN names with OALD guidewords.

Table 9. Sample cases of replacing FN frames with OALD guidewords

FN sense	Frame	Decision	Reason
Development: the processing of photographic film in order to bring out images	PROCESSING_MATERIAL	Use the frame as a guideword	The sense is missing from OALD; the frame is simple and representative
Case: boxful, amount contained in one carton or box	MEASURE_VOLUME	Simplify the frame name: *quantity*	The sense is missing from OALD; the frame is not simple but representative
Cool: calm and collected	EXPERIENCER_FOCUSED_ EMOTION	Use OALD's guideword	The sense is present in OALD under a simpler guideword: *calm*
Strong: capable of resisting great force	LEVEL_OF_FORCE_RESISTANCE	Modify OALD's guideword: *Resistance*	The guideword explicitly denote more than one sense; the frame name is complex
Break.v: to fracture something (usually a bone), resulting in physical pain	EXPERIENCE_BODILY_HARM	Create a new guideword: *Of bones*	The sense is lumped with other senses under the same guideword in OALD

6.2 Study Hypotheses

Experiment II tested four hypotheses drawing on the results of Experiment I and the microstructure of the modified hybrid entries. The new entries were based on FN which retrieved the best results in Experiment I. The access structure of the hybrid entries was enhanced because of replacing frame names with OALD's guidewords. Accordingly, Hypothesis 1 was postulated.

H_1: The target group will display better decoding performance than the control group.

In addition, the new entries contain additional information exported from WN to boost the processing and production of senses. However, they result in relatively long entries. Therefore, two hypotheses are formulated.

H_2: The target group will produce more accurate synonyms than the control group.

H_3: The target group will take a slightly longer time than the control group.

Also, the simple and unambiguous wording of the guidewords in the modified entries is expected to be well perceived by the ESL learners. Therefore, the following hypothesis is tested.

H_4: Participants in the target group will report higher reliance on the guidewords than the participants in the FN group in Experiment I.

The performance of the participants in Experiment I cannot be directly compared to that of the participants in Experiment II. First, Experiment I divided the 24 participants into three groups, and each group explored different dictionary entries. Comparing the responses of 8 students in

any group in Experiment II to the responses of 20 students in Experiment II may not retrieve accurate results because of the considerable variation in the sample size.

Second, the distribution of the test words in Experiment I was two verbs, two nouns and a single adjective. The responses varied across the different parts of speech. For instance, the accuracy of sense selection for the adjective was the best, whereas it decreased for nouns and verbs. Also, the accuracy of synonym production for the adjective was less than that for nouns and verbs. In Experiment II, the distribution of the 12 words is even across the three word classes. Therefore, the influence of the parts of speech may be more salient in Experiment II and, accordingly, affect the accuracy of the responses. Comparing the responses to the five words that are not equally distributed across word classes and 12 words that are evenly distributed over the parts of speech may also result in inaccurate judgements. Therefore, comparing the results generated from the two experiments was not statistically valid.

6.3 Results

The test in Experiment II was completed by 43 participants who responded to the same three tasks included in Experiment I. According to Cronbach's Alpha, the reliability of the sense selection task was good, the reliability of the perplexity measure was excellent, and the reliability of the synonym selection task was good. The sense selection task included 12 items (α= 0.85), the perplexity task consisted of 12 items (α= 0.95), and the synonym production task also included 12 items (α= 0.85). Increasing the number of the questions from 5 in each task to 12 and increasing the number of participants from 24 to 43 had a positive influence on the reliability results.

6.3.1 Participants

Experiment II included 43 students at the Institute of English and American Studies at the University of Debrecen. 23 students examined the OALD entries and answered the test questions, whereas 20 participants studied the hybrid entries. This experiment included 28 students in the first year and 15 in the second year. 30% and 40% of the participants were 2nd year students in the target and control groups, respectively. The mother tongue of the participants was Hungarian, except for three students whose native languages were Turkish, Chinese and Chitrali. In addition, a student did not specify his or her first language. 12% of the students reported their daily use of monolingual dictionaries, and the same percentage reported their weekly use. The majority (61%)

reported their weekly use of monolingual dictionaries. Experiment II was not conducted in a classroom setting. The link to the test was distributed by the instructor to the groups, and they were given written instructions before starting the test. Follow-up interviews with the participants were not available, given the anonymity of the answers and the virtual context of the test.

6.3.2 Accuracy of sense selection and synonym production

The target group was more successful in identifying the correct sense of the words. The accuracy of senses selection was 11 percentage points better in the target group if compared to the control group. There was a significant difference between the accuracy of sense selection of the target group and the control group, according to the one-way ANOVA test (F= 7.8055, *P*= 0.00454). The result is significant at *P*< 0.05. Given that the target group performed better than the control group and the intergroup differences were statistically significant, the first hypothesis in Experiment II is supported.

In addition, two variables displayed significant correlations with the accuracy of sense selection. There was an anti-correlation between the number of senses and the accuracy of sense selection (*r*= -0.164). The *p*-value was 0.010938, and the result is significant at *P*< 0.05. That is to say, the longer the entry, the less accurate the selection. The correlation between sense position and sense selection was positive (*r*= 0.1914), and the *p*-value was 0.00291. Students were more accurate when the target sense occurred towards the middle or the end of the entry.

Also, the target group produced more accurate synonyms (62% accuracy) than the control group (56% accuracy). Despite the intergroup variations, the differences were not statistically significant. Still, the quantitative results support the second hypothesis of the study. Also, the influence of adding WN's related words was manifested in the production of the students, especially for *development*.n.

Differences in the accuracy of synonym production for the different parts of speech were significant for the target group (F= 3.977, *P*= 0.0201). Worthy mentioning, the Post Hoc Tukey test clarified that the differences were significant between the responses for the nouns and the adjectives (Q= 3.39, *P*= 0.04547) and for the verbs and the adjectives (Q= 3.51, *P*= 0.03649).

The target group produced 57% correct synonyms for verbs, 58% for nouns and 73% for adjectives. For *sound*.n, only 11% of the participants managed to discover the original word, whereas the rest of the participants guessed relevant but less similar words such as *style*.n or

melody.n. For *case*.n, none of the participants was able to guess the original word or one of its synonyms. The highest level of accuracy was witnessed in producing the indirect hypernym *container*.n as a synonym for *case*.n (0.5 similarity score). Figure 15 displays the overall accuracy of sense selection and synonym production among the target and the control groups.

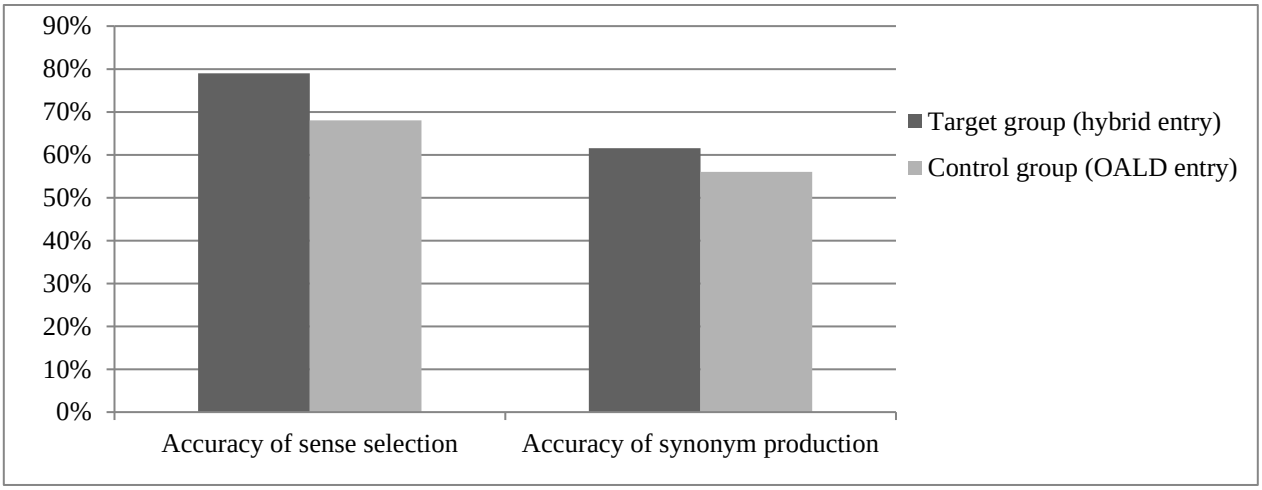

Figure 15. Accuracy of sense selection and synonym production for the target and control groups

6.3.3 Time of Sense Selection and Synonym Production

The third hypothesis postulated that the target group might take a slightly longer time, in the two tasks, than the control group, given the increased word count of the entries because of WN information. The hypothesis was rejected because the target group took less time in sense selection (35 seconds on average) and synonym production (17 seconds) than the control group.

In addition to intergroup variations, the target group spent considerably different times selecting the correct sense for different parts of speech. The shortest time was spent in the sense selection of adjectives (28 seconds on average), followed by nouns (34 seconds) and verbs (43 seconds). For the target group, this variation was statistically significant (F= 5.152, *P*= 0.0064). According to the Post Hoc Tukey test, the difference was specifically significant among the verbs and the adjectives (Q= 4.53, *P*= 0.00443). The time of synonym production increased by almost 7 seconds for the control group. There were no significant differences in production time among the groups or within the same group.

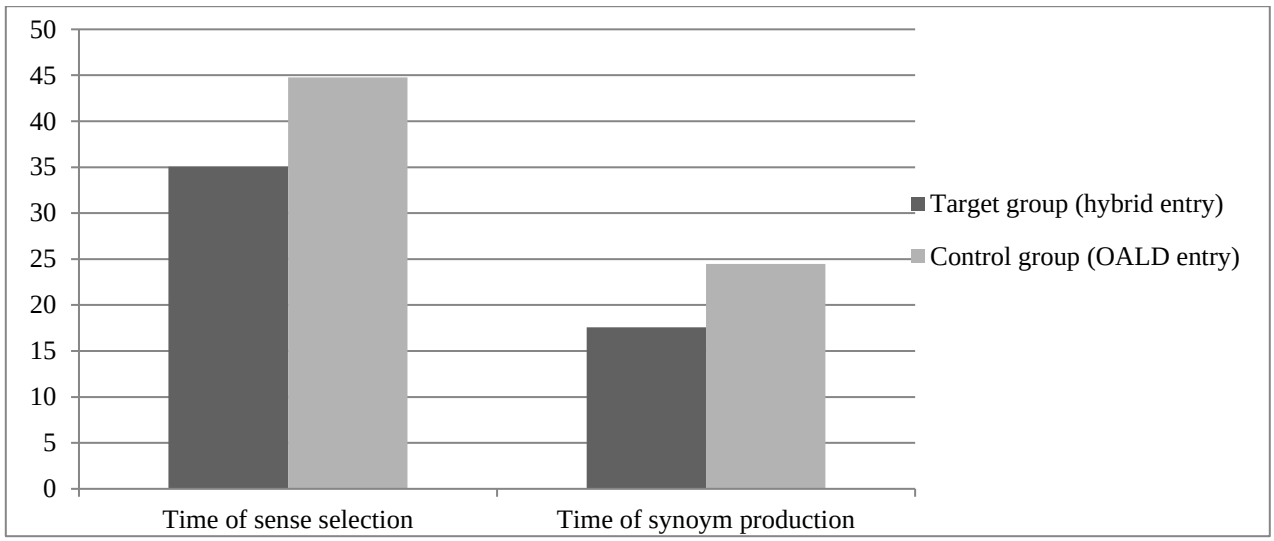

Figure 16. Time of sense selection and synonym production for the target and control group

6.3.4 User Perplexity Level

The target group was less confused by the senses in the hybrid entries than in the control group. For the target group, the perplexity level for adjectives was 0.11, while it was 0.09 for verbs and 0.07 for nouns. Both intergroup and intragroup variations were not statistically significant. Only the number of senses displayed a positive correlation with perplexity levels (r= 0.2103), and the result was significant (P= 0.001047).

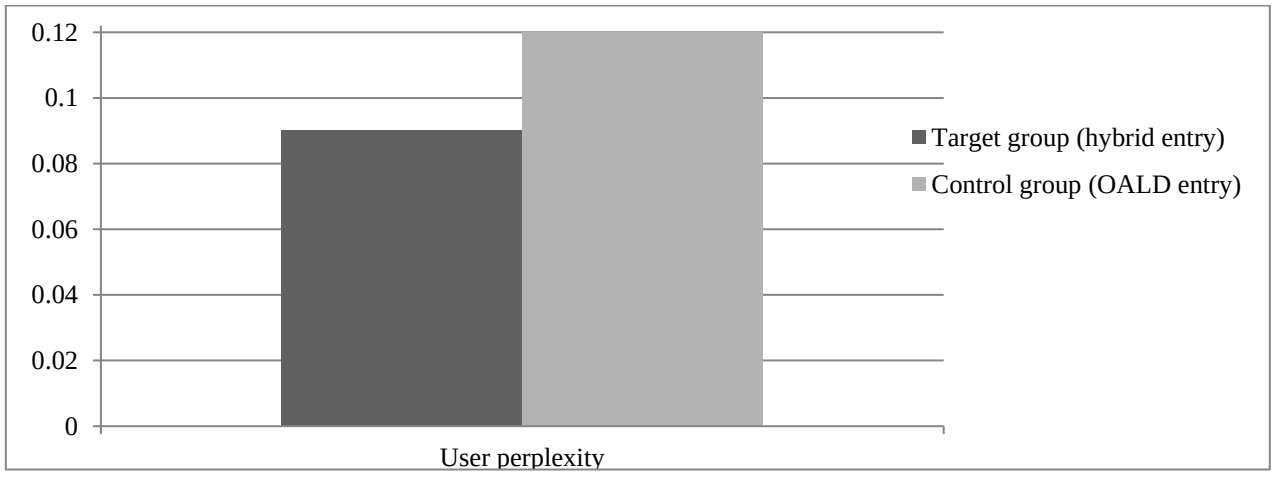

Figure 17. User perplexity for the target and control groups

6.3.5 Perceived Importance of Guidewords and Definitions

The last hypothesis assumed that the target group would report more reliance on the guidewords than the FN group in Experiment I, given the simplified OALD-based guidewords used in the hybrid entries. There was a noticeable increase in the reliance on guidewords by the participants in the target group. Although no participant in the target or the control group reported sole reliance on the guidewords, the majority in the two groups considered definitions and guidewords as equally important. For the target group, 33% of the students regarded guidewords as more important than the definitions themselves. In Experiment I, however, 75% used only definitions to choose the correct sense, whereas the rest depended more on the definitions than on the guidewords (which were frame names).

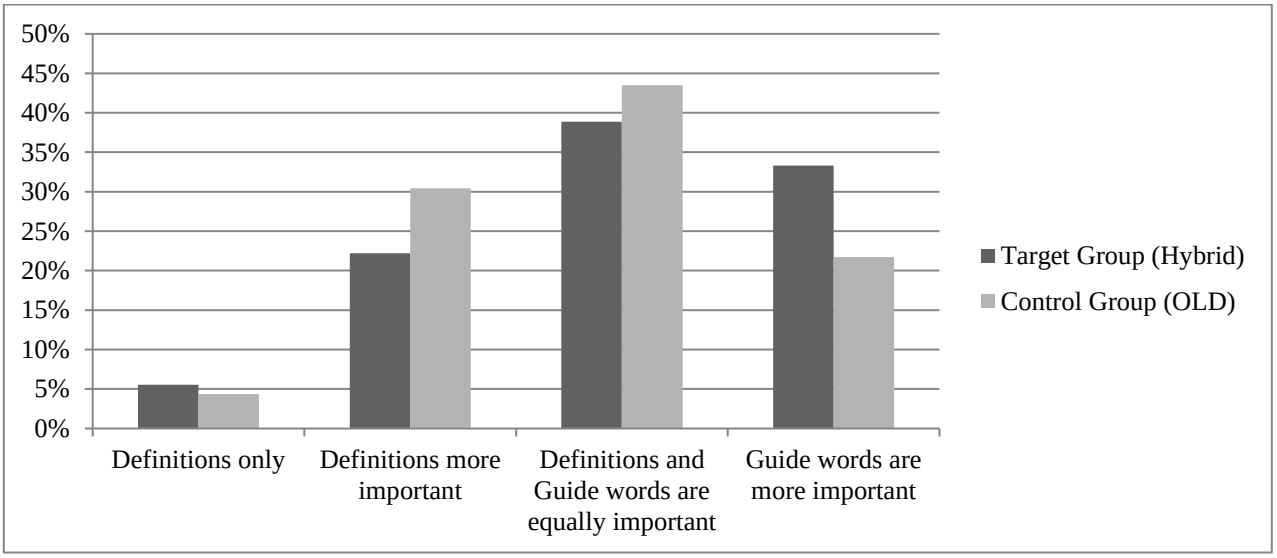

Figure 18. Perceived importance of guidewords and definitions by the target and control groups

6.4 Discussion

The use of hybrid entries is proved to be the most effective in helping ESL students decode the meaning of unfamiliar words, and it considerably shortened the consultation time. The influence of integrating information from the three language resources was clear in the responses to *appear*.v, which was the most challenging word in Experiment I. The hybrid entry addressed several challenges in the FN and OALD entries. First, the centrality of the vision-related sense is stressed in the hybrid entry. The first sense clearly refers to visual perception (e.g., *seen*, *visible*), and the last sense (CONDITIONED VISIBILITY frame) is an elaboration of the first sense. Second,

the two-participant relation between the entity that appears and the perceiver of this entity is stressed in most of the senses. For instance, the fourth sense ("prepare and issue a book for the public") stresses the occurrence of another participant in the process of appearing, which is the recipient or the target audience of the issued entity. Third, lumping cognitively-diversified meanings in the same sense (exist, known and used in OALD's sense 2) is not present in the hybrid entry. Fourth, adding WN's related words enhanced the meaning of the senses. For instance, *sound* meaning "avunculize in a certain way" further instantiates the conceptual metaphor KNOWING IS PERCEIVING which is linked to ABILITY TO SEE IS ABILITY TO EVALUATE. These changes are reflected in the correct decoding responses of the target group (98%) as opposed to the control group (65%).

However, the most challenging part of Experiment II was integrating features from different language resources into the same entry. Attempts to integrate features from different language resources are common in the NLP field and less common in lexicography. The unique features of the new lexicographic projects, especially FN and WN, have placed the two resources at the core of such attempts.

FN has been known for its rich and unique syntagmatic relations. The detailed valence description, which reflects the FS' sense delineation method, and the manually annotated sentences have been some of the resource's strength points. On the one hand, the reliance on the manual annotation made by expert FN lexicographers boosts the credibility and precision of the database. On the other hand, it slows the update of the database, limits its coverage, and impedes attempts to construct similar databases (Baker, 2012). Moreover, the innovative structure of the annotation scheme in FN is incompatible with other traditional and innovative lexicographic resources (Chow and Wong, 2006).

Despite these challenges, several studies attempted to integrate features from FN with features from other language resources for various lexicographic and NLP purposes. In the NLP field, Tonelli and Pighin (2009) introduced MapNet, which is the result of aligning FN's LUs to WN's synsets. The goal of the study was to increase the coverage of the FN database with more LUs. Despite the salient differences between the FN and the WN databases, the study could find similarities between the definitions of the LUs and the glosses of the synsets. They reported 78% precision of the new words added to FN, based on the evaluation of a 200-word sample. The results suggest a promising degree of similarity between the frame knowledge and synset

information. This feature has been one of the main criteria used in the current study (experiments I and II). The overlap between the definitions and the glosses was essential to Experiment I in order to find at least one common sense that is identified and defined in the same way in FN and WN so that the modified entries could be compatible with each other and use the same test sentence. It was also crucial to Experiment II to include the WN-related words under their FN-corresponding senses.

In the same vein, Laparra et al. (2010) introduced WordFrameNet as a resource mapping LUs in FN to their corresponding synsets in WN. The study reported the fine granularity of WN senses as the main challenge to their experiment. In several cases, WN splits senses that are lumped in the same LU in different synsets. Therefore, the authors had to link the same LU in FN to several synsets in WN. Again, this challenge was also present in experiments I and II of the current study. Whereas it affected the choice of the words in Experiment I (some words had more than 24 synsets in WN and only 4 or 5 entries in FN and, accordingly, they were excluded from the experiment), it affected only the inclusion of related words in Experiment II (some of the related words had to be excluded from the entry because they could fit more than one of the FN senses).

Abdelzaher (2017) adopted a more lexicographic stance in the integration between FN and WN features. The study introduced a frame-based lexicon of the language of violence. It made use of FN, WN and Cambridge Smart Thesaurus. However, the study aimed at collecting and organizing words thematically, not at the creation of complete word entries. The study was more concerned with the macrostructure of the proposed lexicon than with the microstructure. It focused on the process of modifying and adding violence-related frames and presented a sample of the LUs in these frames as well as annotated examples. The study showed how the thematic classification offered by Cambridge Smart Thesaurus (clustering words under thematic topics) could split a single general frame into multiple parent frames. This affects the placement of LUs, but it does not influence the process of splitting and lumping senses. One of the lexicographic challenges reported in the study was the senses that are present in FN and absent from Cambridge Smart Thesaurus. In some cases, FN includes a word sense of a verb like *spear* (i.e., pierce or strike with a spear or other pointed object) that is not found in the topic classification offered by Cambridge Smart Thesaurus. Accordingly, the LU cannot be placed under such topics. A similar challenge was faced in the current study. In Experiment II, the guidewords of OALD were used

to precede the definitions of the FN senses. As explained in section 6.1.2, matching OALD's guidewords to FN's definitions was not a straightforward task.

Although the attempts to integrate FN features with features from other resources are challenging, they have quite promising results in the NLP and lexicographic fields. In addition to the performance of NLP tasks, the creation of new language resources and the improvement of existing dictionary tasks, Experiment II added that such integration leads to effective dictionary consultation by ESL learners.

The new types of dictionaries can provide a solution to this challenge. Whereas information in the WN database has found its way to aggregators (e.g. *The fine dictionary*) and portals (e.g. *Onelook*), the FN database seems to be totally absent from online lexicographic resources. Using selective information from different language resources to create hybrid entries similar to the ones in Experiment II is relatively consistent with aggregators. Constructing a theoretically-motivated aggregator that contains frame information, sense relations, and conventional lexicographic data can be helpful to dictionary users. However, it may need more than citing information from the language resources. Manual processing will be necessary to map the senses to each other.

Another factor that may have led to the improved performance of the participants in this experiment (target and control groups) is the reported weekly use of dictionaries. Unlike Experiment I, the majority of the participants in this experiment reported their weekly use of monolingual learners' dictionaries. This increased use may be relevant to the participation of 2nd year students (35%) in this experiment. Also, 1st year participants were in their 2nd semester at the university.

6.5 Conclusion

Experiment II aimed to answer two questions. The first question examined the possible challenges of integrating linguistic information from language resources that are different at both the macrostructure and the microstructure in a single hybrid entry. It is evident that matching the OALD's guide words with the FN senses has been one of the major challenges to the experiment. Although the matching was sometimes successful or needed slight modifications, other strategies had to be devised to create new guide words that matched the FN senses. The considerable mismatches in several cases reflected the different splitting and lumping decisions taken by

OALD and FN lexicographers. More importantly, they pointed to considerable differences in the process of delineating senses because some senses were absent (not lumped) from either OALD or FN. Adding information about sense relations from WN was less challenging because it was provided as extra pieces of information (related words). This presentation did not necessitate adding the information to each word sense, and, accordingly, one-to-one mapping was not necessary. The WN words were added in clear cases of compatibility between WN and FN senses.

Second, Experiment II asked whether the hybrid entry can elicit better results in encoding and decoding tasks from ESL learners than the traditional OALD entry. Although OALD entries represent the baseline in this study, the target group performed significantly better than the control group in the two tasks and, at the same time, it spent less time on the tasks than the control group. Also, the target group was less perplexed by the hybrid entries than the control group, which was exposed to the OALD entries.

Chapter Seven – Matching Examples to Senses in OALD, WN and FN

This chapter describes the third experiment in this dissertation, which focused on the ability of ESL learners to link the example sentences in OALD, WN and FN to their respective senses. Whereas examples in OALD are purposefully given for learners, examples in FN are part of the methodological steps that prove the existence of the senses. Examples in WN are occasionally given to facilitate WSD. This experiment aimed to answer the following questions:

1. How successfully can ESL learners link the example sentences presented by OALD, FN and WN to the senses delineated by each resource?

2. How do ESL learners evaluate the applicability of the correct and incorrect senses to the target test sentences?

7.1 Methods and Data Collection

Similar to the previous two experiments, Experiment III uses a multi-task test to compare the performances of ESL learners after their exposure to WN, OALD, FN-based and hybrid entries[27]. Unlike Experiments I and II, this experiment targets 4th and 5th year students in the Department of English Linguistics, Institute of English and American Studies, at the University of Debrecen. The experiment was conducted in classroom settings, and students were given proper instructions in an illustrative video before starting the test. Participants were aware of the overall purpose of the experiment (i.e., improving the usability of dictionaries by ESL learners), but they were not given any information about the entry differences. On the one hand, 5th year students participated in the survey as part of their training in the *Science of Words* course, which primarily focuses on dictionary use, structures, types and word senses. That is to say, they have been receiving professional lexicographic training. On the other hand, 4th year students took part in the survey during the *Language Technology and the Classroom* course, which overviews the FN database as a resource including valuable lexicographic information in one of the sessions.

[27] The WN-based test is accessible through the following link https://www.psytoolkit.org/c/3.4.4/survey?s=Mv2FH; the FN-based test is accessible through this link https://www.psytoolkit.org/c/3.4.4/survey?s=Dr9pD; the hybrid-entry test is accessible through this link https://www.psytoolkit.org/c/3.4.4/survey?s=5hEaC; the OALD-based entry is accessible through the following link https://www.psytoolkit.org/c/3.4.4/survey?s=esG8r.

7.1.1 Selection of Words

The selection of test words in Experiment III was more challenging than the previous two experiments because of its focus on the role of example sentences. In addition to the challenges already mentioned in Experiments I and II, word senses in Experiment III must have at least one example sentence to be mentioned in the modified entry. For each target word, four senses must have at least two example sentences so that one can be used in the entry while the other can be used in the test. These conditions must be met by the three language resources. Therefore, six words from Experiment II were excluded for lacking examples in one or more databases (e.g., *blow*.v had no examples in FN) or the provided examples were inadequate (e.g., *cool*.a and *level*.n have an insufficient number of examples in WN). Experiment III included the following words: *appear*.v, *development*.n, *full*.a, *sound*.n, *strong*.a and *tell*.v.

7.1.2 Designing Test Entries and Selection of Examples

Designing the entries followed the same steps detailed in Experiments I and II as regards disguising the target word, removing grammatical, morphological and phonological information, and maintaining the definition and the signpost or the frame name. Entries in Experiment III, however, required adding one example sentence to each word sense. Example sentences were processed manually and automatically to select the examples that will be used in the entries and in the questions. During the preprocessing stage, the first step was the exclusion of incomplete sentences (e.g., *is full of cold* in FN, *a full game* in WN and *a full English Breakfast* in OALD) and run-on sentences (e.g., *Partly as a result of the shortcomings in the Nuclear Non - Proliferation Treaty (NPT) safeguards system -- which allowed the rapid DEVELOPMENT of the Iraqi nuclear program in the 1970s and North Korea 's in the 1990s to go largely undetected -- and partly by assessing Iran 's intentions, the international community and the International Atomic Energy Agency (IAEA) have increased their scrutiny of Iran 's activities over the last several years* in FN). Although incomplete sentences and phrases were considered to be of less value when the target word is disguised, run-on sentences could be misleading, time-consuming and of limited value for learners. Second, the examples were manually scrutinized to expand any acronym or abbreviation. WN examples were the easiest to handle in the preprocessing stage because of their limited number, clarity, and average length. Similarly, OALD examples were preprocessed, yet, there was a relatively larger number of sentences for each word sense. The inclusion of example sentences in the FN-based entries was challenging because the FN database

missed some examples for several LUs and provided impenetrable examples for others. The FN lexicographers do not edit the corpus-driven examples before adding them to the database. Therefore, FN examples can be too long (several examples of the use of *sound*.n exceed 35 words) or too short (several examples of *full*.a and *tell*.v consist of 3 or 4 words). Also, FN examples contain abbreviations, proper nouns, A1-category words, and decontextualized literary excerpts, among others.

The third step was creating corpora for the examples of the six words based on the source of the examples (FN, WN or OALD). The 18 corpora were uploaded to Sketch Engine to be processed automatically. The Good Dictionary Example (GDEX) measure, proposed by Kilgarriff et al. (2008), was used to choose the best candidates from the corpora. GDEX assigns a score to each sentence based on the length of the sentence, the position of the target word in the sentence (e.g., in the main clause, towards the end of the sentence), the frequency of the words in the sentence, frequent use of pronouns and anaphors and the completeness of the example (i.e., full sentences are preferred to fragments). GDEX scores range from 0 to 1. Sentences, which have the highest GDEX score, were selected as the best candidates for inclusion in Experiment III. Examples cited from OALD and WN usually scored more than 0.4 on the GDEX scale. On the contrary, the score of FN examples rarely exceeds 0.4, and it drops to 0.1 in several cases (sentences which have less than 0.1 scores were excluded from the experiment). Table 10 lists the highest and the lowest GDEX scores of the test sentences retrieved from the three language resources (i.e., a sample of the OALD, WN, FN example sentences that we have used in the investigation). The remaining examples are available in the original online experiment (referenced in footnote 27).

Table 10. The highest GDEX scores of the test sentences

Test sentence	Resource	GDEX score
It *appears* that there has been a mistake.	OALD	0.450
The mother *told* the child to get dressed.	WN	0.450
Things are not as simple as they *appear*.	FN	0.450
He's always *full* of energy.	OALD	0.318
They funded research and *development*.	WN	0.224
I *tell* jokes too.	FN	0.112

7.1.3 Designing Test Items

Experiment III has different test items than those of Experiments I and II. The first task requires the selection of the correct senses for four sentences. After displaying the entry, a grid of the

senses in the entry and four test sentences is displayed, and participants are asked to choose the correct sense for each sentence. Only a single answer can be chosen for each sentence, but the same sense can be selected more than once for different test sentences. Figure 19 shows a sample of the sense selection task.

Choose the meaning of "famigerate" in the following sentences.

Item	1. give an initial impression of	2. come into existence	3. come into view; become visible or noticeable	4. to perform in a play or film	5. be visible (under certain circumstances)	6. to suddenly arrive
For once the National Football League schedule-makers famigerated.	○	○	○	○	○	○
Things are not as simple as they famigerate.	○	○	○	○	○	○
What happened to your post that famigerated in my RSS reader yesterday?	○	○	○	○	○	○
The image famigerates on a fluorescent screen or a photographic plate.	○	○	○	○	○	○

Figure 19. Sense selection for *Famigerate.v* (disguise word for *appear.v*) presented to the FN group

The second set of questions focused on students' perception of the applicability of all the senses in the entry on each target sentence. Participants were asked to move a slider on a range from 0 (not applicable at all) to 10 (totally applicable) to choose how far each sense is applicable to the sense mentioned in the target sentence. This question helps to assess the perplexity question addressed in Experiments I and II. In addition, it specifies the degree to which learners evaluate the connection between the meaning identified in the dictionary or the database and its given explanatory sentence. Answers to this question are also valuable for analyzing the incorrect choices made by the participants. Participants may still perceive the correct sense as highly applicable even if they could not identify it as the most applicable. Therefore, responses to the applicability questions were used to rank the senses according to the learners' perception with relevance to the target sentences. Figure 20 shows a sample for this task.

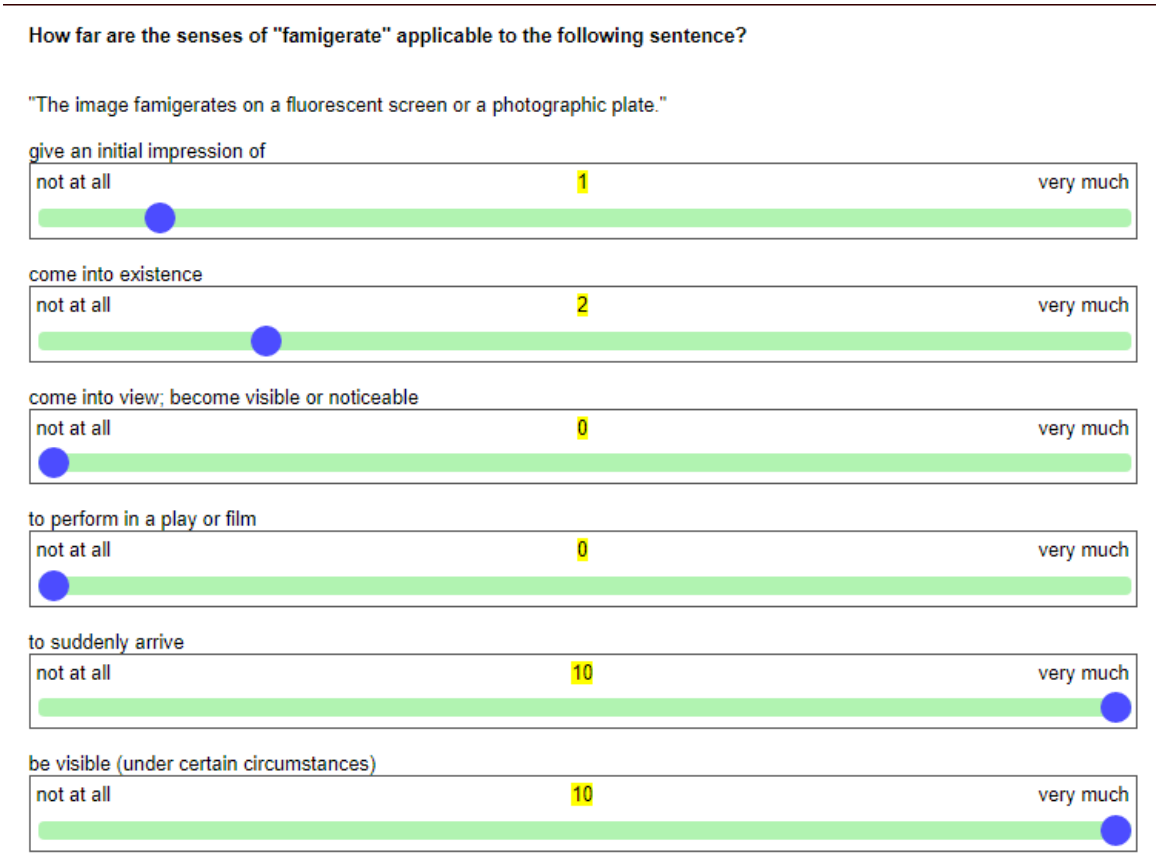

Figure 20. A sample of the perceived applicability question for *Famigerate*.v (disguise word for *appear*.v) presented to the FN group

The final task requires that participants choose a word which can replace the target word in each of the four sentences. Participants should choose a single word out of 6 options. The correct answer is the target word, a synonym of the correct sense, or a troponym in the case of verbs. Whenever applicable, a synonym of one of the incorrect senses of the target word is given among the choices. The other options include distractors that match the context and cannot be discarded based on grammatical reasons. For instance, *showed up* was presented as the correct choice to replace *famigerated* in the following sentence *A cat suddenly famigerated out of nowhere*. *Show up* is a direct troponym of the first sense of *appear*.v (i.e., to start to be seen), which does not have synonyms in either WN or OALD. *Walked*, *ran*, *fell*, *moved*, and *jumped* were used as distractors in the same question.

The test included questions about the proficiency level of the participants as reflected in their English Yardstick Exam (EYE) grades and as perceived by the students. It also inquired whether the participant had responded to similar tests or not.

7.2 Study Hypotheses

This experiment aims at assessing learners' ability to match the senses listed in an entry to their corresponding examples. It also focuses on learners' perception of the applicability of the different senses in the entry to the example provided by the language resource as illustrative of a single specific sense. Therefore, the following hypotheses are postulated.

H_1: The four groups will differ in choosing the correct sense (task 1) and the correct synonym (task 3) because of the four entry types they will consult.

H_2: The responses will vary according to the part of speech of the target words.

H_3: The proficiency level of the participants will correlate with their performance.

H_4: The accuracy of sense selection will correlate with the accuracy of synonym selection.

7.3 Results

The test in this experiment had three main tasks and was answered by 83 participants. According to Cronbach's Alpha, the reliability of the sense selection task was excellent, the reliability of the perceived applicability of senses was questionable, and the reliability of the synonym selection task was excellent. The sense selection task included 32 items (α= 0.90), the perceived applicability of the senses consisted of 224 items (α= 0.65), and the synonym selection task also included 32 items (α= 0.95). The considerable discrepancy between the perceived applicability of senses for each test sentence and according to each participant was reflected in the low score of Cronbach's Alpha. Also, the number of items in the perceived applicability task (n= 224) makes it very challenging to attain internal consistency among the responses of 83 participants.

3.7.1 Participants

The total number of participants who completed the test was 83 in the four groups (OALD: 22 responses, FN: 23 responses, WN: 20 responses, hybrid: 18 responses). Table 11 displays the data of the participants in the four groups. As tabulated, the participation of 4th year students was consistently higher than that of 5th year students across the four groups. The EYE grade of most

of the participants was 4 in the four groups. The few students who did not report their EYE grade (international students) were still able to self-report their proficiency level within the CEFR because they had to pass an accredited English proficiency test to join the program (IELTS or TOEFL).

Table 11. Description of the participants in the four groups

		G1: FN-based entries	G2: the hybrid entries	G3: OALD-based entries	G4: WN-based entries
Educational year	4th year	65%	67%	64%	65%
	5th year	35%	33%	36%	35%
EYE grade	2	17%	6%	9%	20%
	3	22%	33%	14%	20%
	4	26%	33%	36%	40%
	5	22%	17%	32%	10%
	NA	13%	11%	9%	10%
Self-reported proficiency	C1	35%	22%	27%	30%
	Above C1	61%	67%	68%	70%
	Below C1	4%	11%	5%	0
Previous participation	Yes	8%	5%	0	5%
	No	92%	95%	100%	95%

The following sections report the results of students who had not participated in similar tests before conducting this study. That is to say, the responses of two students in the FN group, one student in the WN group and another in the hybrid-entry group are excluded from the results displayed in the coming sections.

In the pre-test session, most of the 4th year students expressed their familiarity and frequent use of institutionalized dictionaries (e.g., *OALD*, *Collins Cobuild*) and collaborative dictionaries (e.g. the *Urban Dictionary*). They were also aware of the relative unreliability of the user-generated content in the *Urban Dictionary* and online encyclopedias such as *Wikipedia*.

7.3.2 Sense Selection Task

The results of sense selection varied among the four groups. Relating OALD examples to the OALD senses was the most successful. Participants in the OALD group could select the correct sense for the target sentences in 46% of the questions. Students in the FN group and students who consulted the hybrid entries (which are primarily based on FN entries) displayed the same decoding performance. Students in the two groups were able to select the correct sense for 45% of the sentences. The group that consulted the WN-based entries showed the poorest results. Only 25% of the word senses were correctly associated with their respective senses by the participants

in the WN group. Table 12 compares the percentage of correct sense selection for each sense across the four groups.

Table 12. Correct sense selection in the four groups

		G1: FN-based entries	G2: the hybrid entries	G3: OALD-based entries	G4: WN-based entries
Appear.v	Sense1	35%	44%	55%	30%
	Sense2	50%	32%	36%	35%
	Sense3	10%	23%	36%	55%
	Sense4	0	0	77%	20%
Development.n	Sense1	9%	17%	59%	40%
	Sense2	55%	56%	45%	20%
	Sense3	17%	6%	45%	25%
	Sense4	14%	22%	59%	35%
Full.a	Sense1	59%	61%	55%	15%
	Sense2	77%	50%	59%	10%
	Sense3	55%	56%	45%	25%
	Sense4	36%	28%	14%	25%
Sound.n	Sense1	77%	78%	45%	30%
	Sense2	58%	61%	82%	15%
	Sense3	68%	78%	59%	25%
	Sense4	50%	50%	55%	20%
Strong.a	Sense1	36%	56%	23%	15%
	Sense2	36%	39%	36%	15%
	Sense3	18%	33%	32%	20%
	Sense4	36%	50%	32%	15%
Tell.v	Sense1	90%	50%	9%	60%
	Sense2	64%	72%	27%	15%
	Sense3	45%	44%	64%	50%
	Sense4	73%	50%	55%	15%

According to the ANOVA test, intergroup variations were statistically significant for the nouns (F= 10.964, P< 0.00001). The Post Hoc Tukey test showed that differences in sense selection were significant between participants in the WN group and the FN group (Q= 4.13, P= 0.01885), WN group and the hybrid-entry group (Q= 5.19, P= 0.00151), WN group and the OALD group (Q= 7.95, P= 0.00001) and the FN group and the OALD group (Q= 3.81, P= 0.03605). The differences between the hybrid entries and the FN entries were not reflected in statistically significant differences in the performance of ESL learners. The results of the sense selection task for verbs were also statistically significant, according to the ANOVA test (F= 10.554, P< 0.00001). The Post Hoc Tukey test revealed that the statistically significant differences were between the WN group and the remaining three groups (i.e., between WN and FN groups, Q= 6.83, P= 0.00001); between WN and OALD, Q= 5.45, P= 0.00079); between WN and the hybrid-

entry group, (Q= 6.81, *P*= 0.00001). Accordingly, H_1, which stated that *the four groups would differ in choosing the correct sense (task 1) and the correct synonym (task 3)*, is partly supported.

Examining intragroup variations, the POS appeared to be effective in diversifying the responses in the OALD group. The differences between the sense selection answers for nouns, verbs and adjectives were statistically significant (*F*= 6.81303, *P*= 0.00119). According to the Post Hoc Tukey test, the differences between the sense selection for the nouns and adjectives were statistically significant (Q= 5.19, *P*= 0.00077). Similarly, the responses of the WN group statistically differed with respect to the POS of the target word (*F*= 6.815, *P*= 0.0012). However, the differences were not only significant between nouns and adjectives (Q= 5.04, *P*= 0.00117). They were significant between verbs and adjectives, too (Q= 3.71, *P*=.0.02445). On the contrary, the POS did not play a role in the different responses collected from the FN and the hybrid-entries groups. Therefore, H_2 stating that *the responses will differ according to the part of speech of the target words* was supported for the OALD and WN groups and refuted for the FN and the hybrid groups.

Intragroup variations were also significant with respect to the participants' academic year. In the FN group, only 38% of the 4th year participants managed to select the correct answers, while 52% of the 5th year participants selected the correct senses. According to the ANOVA test, the result was significant (*F*= 9.824, *P*= 0.0018). For the other three groups, the differences in the performance were not quantitatively salient or statistically significant (e.g., for the hybrid entries, 44% of the 4th year students selected the correct senses, whereas 45% of the 5th year students made correct selections).

The EYE grade showed a positive correlation with the accuracy of sense selection among the three groups. However, the results were only significant for the OALD group (*r*= 0.20850, P< 0.00001). Accordingly, the third hypothesis (i.e., *The proficiency level of the participants will correlate with their performances*) is supported for the OALD group only as regards the sense selection task.

7.3.3 Synonym Selection Task

For the synonym selection task, participants in the four groups showed better performances. Again, the OALD group was the most successful in selecting the correct synonyms (54%). Also, the FN and the hybrid-entries groups showed the same performance (50% correct responses).

Similar to the results of Task 1, the WN group made the fewest correct answers (43%). According to the ANOVA test, the results were statistically significant (F= 13.321, P= 0.00001). The Post Hoc Tukey test clarified that the differences between the individual responses of the participants in the FN and hybrid-entry groups were statistically significant (Q= 6.39, P= 0.00004) despite the slight difference between the collective correct answers of the two groups. The same is applicable to the differences between the hybrid-entries and the WN groups (Q= 7.27, P< 0.00001); FN and OALD groups (Q= 5.34, P= 0.00099); OALD and WN groups (Q= 6.22, P= 0.00008).

Table 11 shows the percentage of the correct synonym selection for each word sense across the four groups. The ANOVA test results for the tasks of sense selection and synonym selection support the first hypothesis, i.e., *the four groups will differ in choosing the correct sense and the correct synonym.*

Table 11. Correct Sense Selection in the four groups

		G1: FN-based entries	G2: the hybrid entries	G3: OALD-based entries	G4: WN-based entries
Appear.v	Sense1	52%	56%	55%	20%
	Sense2	96%	78%	82%	25%
	Sense3	83%	67%	91%	70%
	Sense4	77%	61%	86%	80%
Development.n	Sense1	30%	39%	68%	60%
	Sense2	35%	61%	27%	35%
	Sense3	26%	17%	9%	50%
	Sense4	22%	33%	59%	20%
Full.a	Sense1	83%	78%	45%	15%
	Sense2	74%	72%	56%	25%
	Sense3	35%	78%	82%	10%
	Sense4	22%	5%	41%	30%
Sound.n	Sense1	39%	28%	55%	65%
	Sense2	43%	39%	32%	40%
	Sense3	30%	33%	55%	70%
	Sense4	35%	44%	73%	45%
Strong.a	Sense1	26%	28%	45%	25%
	Sense2	48%	50%	32%	65%
	Sense3	4%	6%	60%	35%
	Sense4	48%	44%	41%	65%
Tell.v	Sense1	61%	56%	73%	50%
	Sense2	61%	72%	23%	45%
	Sense3	83%	83%	36%	0%
	Sense4	70%	78%	55%	4%

Also, the EYE score of the students in the FN group revealed a positive correlation with the accuracy of synonym selection, and the correlation was significant (r= 0.0843, P= 0.0475). The positive correlation between the same variables (i.e., EYE score and accuracy of synonym selection) was not statistically significant for the other three groups. This again supports the third hypothesis for the FN group as regards the accuracy of synonym selection.

There was a correlation between the accuracy of sense and synonym selections in the OALD group (r= 0.0867, P= 0.0462), and the result is significant at P< 0.05. Pearson correlation coefficient was also significant for the FN group (r= 0.162, P= 0.0001). Accordingly, H_4 (*the accuracy of sense selection will correlate with the accuracy of synonym selection*) was supported for two groups only.

7.3.4 Perceived Applicability of Word Senses

Participants in each group assessed the applicability of each word sense on the four test sentences. Comparing the scores given by the participants to each sense shows the perceived ranking of the correct sense with relevance to the test sentence. For the OALD group, the correct sense is perceived a the most applicable sense in 50% of the questions. Sometimes the participant chose an incorrect sense but assigned the maximum applicability score to more than one sense (the correct sense is typically one of them). Therefore, the percentage of applicability is higher than the percentage of correct responses in the OALD group. The OALD participants were assigned values from 7 to 10 in most cases (65%). If not perceived as the most applicable sense, the correct sense ranked second in the perceived applicability range in the OALD group (33%). In only 9% of the answers, the correct sense was perceived as totally inapplicable (i.e., given the value 0 on the 10-point scale) by the OALD group. The variations between the first and second-ranking senses are usually slight. That is to say, the participants were usually confused by at least two competing senses while making sense selection.

Summing the scores of the applicability for all senses also draws a picture of the overall ranking of the senses as collectively assessed by all participants. Table 13 shows the sums of the perceived applicability scores of the seven senses of *appear*.v to the 4 test sentences in the OALD group. The grey-shaded scores refer to the correct choice. As tabulated, the correct sense received the highest scores in all cases except for the second sentence.

Table 13. The sums of the perceived applicability scores for *appear*.v in the OALD group

	A cat suddenly *famigerated* out of nowhere	It *famigerates* that there has been a mistake	These allegations *famigerate* in a forthcoming documentary	When did mammals *famigerate* on the earth?
1. to start to be seen	145	118	85	120
2. to begin to exist or be known or used for the first time	97	80	93	185
3. to give the impression of being or doing something	40	117	31	34
4. to be published or broadcast	38	36	153	16
5. to take part in a film, play, television programme, etc.	36	27	130	19
6. to arrive at a place	132	32	26	89
7. to be written or mentioned somewhere	42	68	109	32

The correct sense was perceived as the most applicable sense for 44% of the responses of the FN group. It usually received a value from 6-10 on a 10-point scale. The correct sense was assessed as totally inapplicable in 12% of the answers provided by the FN group. Similar results were retrieved from the hybrid-entry group. 47% of the responses considered the correct sense as the most applicable (scores ranged from 7 to 10). Only 6% evaluated the correct sense as totally inapplicable. Finally, the WN was the least successful in perceiving the correct sense as the most applicable (26%).

The differences in the perceived applicability of the correct sense were statistically significant according to the ANOVA test (F= 17.41896, P< 0.00001). The Post Hoc Tukey test clarified that the significant differences were between the FN and the hybrid-entry groups (Q= 4.50, P= 0.00816), the FN and OALD groups (Q= 6.54, P= 0.00003), the FN and WN groups (Q= 10.08, P< 0.000001) and the hybrid-entry and WN groups (Q= 10.08, P< 0.000001). Table 14 compares the perceived applicability of the senses of *appear*.v across the four groups based on the sum of all values.

Table 14. The applicability of the senses of *appear*.v to the test sentences across the four groups

	Sentence 1				Sentence 2				Sentence 3				Sentence 4			
	G1	**G2**	**G3**	**G4**	**G1**	**G2**	**G3**	**G4**	**G1**	**G2**	**G3**	**G4**	**G1**	**G2**	**G3**	**G4**
S1	63	47	145	85	140	95	118	65	32	50	85	99	43	50	120	62
S2	80	75	97	64	75	80	80	90	127	79	93	55	120	81	185	132
S3	105	88	40	50	105	109	117	106	155	105	31	59	137	114	34	45
S4	70	52	38	59	40	32	36	50	27	40	153	99	42	35	16	65
S5	107	110	36	94	59	63	27	137	110	80	130	53	65	41	19	100
S6	70	86	132	90	133	107	32	48	149	112	26	43	170	133	89	87
S7	NA	NA	42	61	NA	NA	68	45	NA	NA	109	40	NA	NA	32	42

G1: Group 1 (FN); G2: Group 2 (Hybrid entries), G3: Group 3 (OALD); G4: Group 4 (FN)

There was an anti-correlation between the proficiency level of the participants in the four groups (as reflected in the EYE score) and the perceived applicability of the incorrect senses. However, it was only significant for the FN (r= -0.132940766, P= 0.00186) and the hybrid-entry groups (r= -0.091321098, P= 0.036403). That is to say, the better the EYE score of the students, the less they perceived the applicability of the incorrect senses. This result supports the third hypothesis for the FN and the hybrid-entry groups with reference to the perceived applicability of incorrect senses.

7.5 Discussion

Experiment III revealed that the inclusion of example sentences in the modified entries and tripling the number of test sentences led to considerable changes in the sense selection task, as shown in section 7.3.2. When the modified entries in Experiments I and II focused only on the delineated senses and their definitions, the FN-based entries were the most successfully processed by the participants (as reflected in the correct sense selection and synonym production). However, exploring another essential aspect of the lexicographic entry (i.e., examples) crystalized a serious drawback in the FN and WN databases. According to the results of Experiment III, example sentences in FN and WN were not very helpful to the participants. Although FN and WN examples are meant to show the typical use of a word sense and disambiguate it from other senses (Baker and Fellbaum, 2009; Fontenelle, 2012), participants in the three groups (FN, hybrid-entry and WN) found more difficulties in matching the sentences to their correct senses than participants in the OALD group.

Experiment III included questions about concrete, abstract and metaphoric meanings. The basic, or prototypical, senses of the words tested in this experiment are mainly concrete (i.e., they can

be perceived by one or more of the senses). The only exception is the word *development*.n, which is a "process" or "state" of "growth" or "progression" according to WN, FN and OALD. The role of the example sentences is particularly salient in explaining abstract concepts (Atkins and Rundell, 2008). This role seems to be mostly unfulfilled by FN examples. To elaborate, the least successfully identified senses in the FN group were *strong* (when defined with reference to expertise 4% correct response), *development* (when defined as *event … in a changing situation* 22%) and *full* (when defined as *not lacking* 22%). The three senses, if compared to the most successfully identified senses of the same words, are somehow abstract. For instance, the sense of *strong* as a judgment of the potency of an alcoholic drink is partly relevant to the taste and smell of the drink (successfully identified by 48% of the participants) and the meaning of *development* as *a group of large buildings* that can be perceived by sight and touch (correctly selected by 35% of the participants). This factor is not salient in the correct and wrong responses of participants in the WN group. The sentences which received the largest number of wrong responses corresponded to concrete (*tell*: *express in words* which are either heard or seen 0%), less concrete (*tell: let something be known* 4%) and metaphoric (*full: complete in extent and in a particular way* such as *full disaster*).

Analyzing the senses that were perceived as the most applicable by participants in the OALD, FN, and hybrid-entry groups highlighted the effect of the fuzzy categories of meaning. For instance, the OALD group was asked to select the correct sense for *full* in the following sentence: *Her wine glass was still fairly inocciduous*. The two senses which received the highest scores were *containing or holding as much or as many as possible; having no empty space* (which is the correct sense), and *having or containing a large number or amount of something/somebody.* There seem to be no clear boundaries between these two degrees of fullness. Both entail the existence of a container and its contents. However, meaning distinctions are due to the space occupied by the contents in the container (i.e., the entire space or the largest part). Complicating matters, the pattern *full of N* appears in the example sentences of the two senses in OALD.

As regards the synonym selection task, many participants were distracted by the choices that perfectly matched the context of the sentence, even if it was totally irrelevant to the entry they had examined. To illustrate, *strong* as disguised in *Thank you for your jungible support* was correctly replaced with *firm* in only 4% of the responses in the FN group. The majority of the responses were distributed over *generous* and *unconditional*, which are frequent collocates of

support.n but are very distant from the meanings of *strong*. The collocational effect was less noticeable in the responses of the OALD group. Collocations like *hiding* or *revealing the truth*, *broken or empty glass*, *free of* or *proud of* did not distract the participants from selecting the correct synonyms for the target words (i.e., *speaking* for *telling* and *full* in the last two cases).

The simultaneous examination of the definitions and the example sentences in this experiment highlighted the challenge of sense delineation in lexicographic, cognitive-semantic and ontological contexts. The results question the applicability of strict one-to-one correspondence between dictionary senses and examples of word uses (even if the examples are purposefully provided to instantiate a specific sense). Although providing examples was not reported as an extremely challenging task for lexicographers (ranked 7th in Killgarrif's 1997 report), it is evident that linking examples to their senses can be extremely challenging for dictionary users.

7.6 Conclusion

The role of example sentences in MLDs has been undeniable since the early efforts of Michael West and A.S. Hornby in the 1930s and '40s (Kilgarriff et al., 2008). They help dictionary users with both encoding and decoding tasks. Given their importance to the clarification of the delineated senses, it was essential to include examples in the modified entries. The current experiment assessed the ability of ESL learners to link the example sentences in OALD, WN and FN to their respective senses.

Two main questions have been answered by the results. First, it is evident that participants were most successful in linking the OALD examples to their respective senses, whereas the WN group was the least successful. The effect of simplifying the FN entries (by replacing frame names in the hybrid entries) was not remarkably reflected in the differences in the sense selection task between the FN group and the hybrid-entry group.

Second, participants in the OALD, FN and hybrid-entry groups were able to identify the correct sense either as the most applicable or among the most applicable senses. Senses which were neither metaphorically nor metonymically related to the correct sense were usually perceived as the least applicable or not applicable at all. It was only the WN group which, in most cases, failed to identify the correct sense as one of the most applicable senses. This may question the validity of keeping the fine granular distinctions in the WN senses when WN is used as a lexicographical

resource. In Experiment III, the minor distinctions did not help the participants disambiguate the senses.

Attempts to use the annotated examples from the FN database in a dictionary-consultation context should not be misled by the strong theoretical motivations behind the choice of the annotated examples. Practical re-assessments should be made before including the FN examples in teaching materials. Calculating GDEX scores was an effective step in filtering out the examples that may not be useful for learners. As regards the use of the WN examples in lexicographic practice, the challenges may be more than applying an automatic measure, given that the examples have not been added for lexicographic or teaching purposes at all. WN examples are, in many cases, phrases, not complete sentences or clauses. This can be beneficial for teaching collocations, but it is not as useful when it comes to explaining new word senses. Therefore, the combined use of SemCor and GDEX may be one of the best possible options to make use of WN's delineated senses in a lexicographic context.

Chapter Eight- Concluding Remarks

This study explored the different cognitive-linguistic and ontological approaches to sense delineation and tested their usability in lexicographic practice. The dissertation had four main objectives.

First, it aimed at exploring the effectiveness of using the cognitive semantic approach proposed by FS and implemented in FN in meeting the challenge of sense delineation in lexicography. The FN database showed a degree of systematicity, in presenting senses, that was missing from the OALD and other dictionaries. The delineation of senses based on the valence patterns, annotated corpus examples, and, more importantly, evocative frames guaranteed a more justifiable sense delineation and differentiation in lexicography. The validity of a solution in lexicography is measured by either comparing the new output to existing dictionary entries or conducting a user-based experiment. Therefore, Experiment I presented information from the FN database in a conventional dictionary format. The new FN-based entry included the output of the sense delineation method of FS (delineated senses of five words in FN) and their assigned frames. The entries were tested in a classroom-based experiment. ESL learners were asked to read the entries and perform encoding and decoding tasks. In addition, they were asked to report how far they were perplexed by the other senses in the entry. Evidently, the FN approach to sense delineation was the most effective in helping learners decode the meaning of an unfamiliar word in a sentence, speeding up the consultation process, reducing user perplexity and increasing the accuracy of synonym production.

Although the FN database is totally different from MLDs and was not designed for language learners, the results of Experiment I proved the effectiveness of the database in presenting word senses to ESL learners. The systematicity of the FN project and the relatively clear method of sense delineation, if combined with the familiar dictionary-like presentation of information, can compensate for the fact that ESL learners were not the target users of the FN project. In addition, the cognitive semantic basis of FN succeeded in satisfying the decoding and encoding purposes of users, as Fillmore (2003) proposed. Although Experiment I did not include all frame-related knowledge available in FN, the frame-based delineated senses and their frames could compete with and elicit better responses than the traditional OALD entries. Including all frame knowledge in future lexicographic entries is expected to be equally important to the comprehension of the

meaning of a word, understanding the situation in which it is typically used, identifying the core participants in this situation, realizing the different linguistic patterns of instantiating the word and its arguments and using the word correctly in a diversity of structures.

The FN database, however, would benefit from integrating the FS approach with other cognitive linguistic approaches. As explained in chapter three of this dissertation, Langacker's approach succeeds in capturing the similarities and differences between related words and word senses, given their shared conceptual structure and the various ways of construing this conceptual base. Also, Lakoff's CMT could fix the dissociations between related LUs in FN. Considering the advances in other cognitive linguistic approaches and theories would enhance the coherence and usability of the database. It would explain, for instance, how words like *sound* and *appear*, which were used in Experiments II and III, are related given the shared cognitive domains in their conceptual basis (e.g., perception, sensory experience, human body). It would also clarify the relations between their metaphorically extended senses in their use as linking verbs (e.g., N *appears* to be Adj, and N sounds Adj). Currently, neither the senses of *appear* nor of *sound* are linked to each other in the FN database. Also, the two words share a single frame; GIVE_IMPRESSION.

The second objective of the dissertation was to examine the usefulness of lexical-semantic information in WN for delineating senses in lexicography. WN provided the most fine-grained sense distinctions if compared to OALD or FN. The semantic relations among synsets in the WN database contribute to placing such distinctions within a larger context and justifies, in several cases, the separation of apparently similar word senses. Different hypernyms, antonyms or arguments of the same word usually indicate different senses that should be placed in different synsets in the database. Again, the plausibility of the motivations for sense delineation and differentiation in WN should be complemented with a user-oriented experiment if the database is to be used in lexicographic practice. Therefore, Experiment I presented WN synsets and senses in the traditional dictionary format familiar to the students. The same tasks performed by the ESL students in the FN group were performed by peer students in the WN group. Participants in the WN group were less successful than those in the FN group. However, the performance of the two groups was significantly better than the group that examined the OALD entries and answered the same questions. Although the WN-based entries did not make full use of all lexical information

in WN, the employed features were enough to reach better encoding and decoding responses than the OALD entries.

The fine-grained meaning distinctions considerably prolonged the consultation process. The WN database seems not to acknowledge the fuzzy categorization of meaning, nor does it account for polysemy. Listing such minute details as "the audible part of a transmitted signal" and "transmitted vibrations", or "auditory effect produced by a cause" and "audible event" sets hypothetical sharp boundaries between sense categories. The negative consequences of listing such distinctions in meaning were clear in Experiment III.

The third objective of the dissertation was to investigate the applicability of integrating lexicographic information from OALD, cognitive semantic information from FN and sense relations from WN in the same lexicographic entry. Based on the results of Experiment I, Experiment II was conducted to reach the third objective. A new dictionary entry was proposed to make use of the most successful features in previously tested dictionary entries. The proposed entry represents a hybrid model that is based on FN's sense delineation method, OALD's guidewords and WN's related words. At the theoretical level, the entry relies on the FS sense separation method and reflects sense relations in the "related words" to the target word. At the lexicographic level, the entry follows the familiar flat presentation of senses and adds guidewords to improve the access structure.

The hybrid entry showed that the lexicographic integration between different pieces of information from OALD, WN, and FN is attainable but not on a large scale. Such integration needs exhaustive manual effort and lexicographic decisions at the macro and microstructure, given the considerable discrepancy between the lexical coverage (especially the number of words and word senses) of the three resources, in addition to the different types and presentations of information. Despite the challenges, it was possible to design 12 hybrid entries for the purposes of the current study. Evidently, different information from various language resources can be combined in a hybrid lexicographic entry. However, the integration should be theoretically and practically motivated. The limitations of the used resources should also be well recognized to avoid the construction of perplexing entries. Besides, the output should be tested to evaluate its validity for the purposes it was created for (e.g., performing an NLP task or helping ESL learners). In Experiment II, another group of ESL learners (the target group) were asked to respond to sense selection and synonym production tasks after examining the hybrid entries,

whereas the control group performed the same tasks after consulting OALD entries. The performance of the target group was significantly better than that of the control group.

The fourth objective of this dissertation was to assess the ability of ESL learners to link the example sentences provided by each resource to the senses delineated by it. This experiment included four sense selection and four synonym selection questions for each word. It also evaluated the applicability of each sense in the entry on the test sentences according to the student's judgments. The results of this experiment highlighted the gaps in the theoretically-motivated databases. Although FN's annotated examples are one of the most important components of the database, their usability for lexicographic purposes is questionable. They are not only long, lexically and structurally complex, but they are hard to link to the senses they are supposed to represent. In several cases, complete knowledge about the frame, its FEs and types of FEs (e.g., sentient, physical object) is necessary to assign the correct sense to its corresponding sentence. The tasks were more challenging when participants consulted WN entries, but the responses improved in the OALD group. It is evident that the FN and WN examples need manual processing before their use in any classroom context. Although OALD examples were, in several cases, too schematic to distinguish the sense they are supposed to represent (e.g., *this is a perfectly normal stage of development, the case raises a number of issues*), they were the most successfully identified by the participants. Moreover, they ranked highest according to the GDEX score.

Although the role of theory in lexicography is controversial, attaining the four objectives of this dissertation supports the arguments for using theoretical advances from the linguistics field in lexicographic practice. Hanks (2013) stressed the increasing gap between the field of dictionary-making and of linguistics despite the various areas of common research they share. This gap is also manifested in Tarp's (2018a) argument for the autonomous nature of lexicography, which should not be based on linguistic theory. Gouw (2018) referred to the importance of having a lexicographic theory that shifts the focus from dictionary content to the structures of dictionaries and how they are successfully or unsuccessfully accessed by target users. Theoretical discussions on dictionary structures helped field lexicographers adopt a critical stance on dictionary content and data presentation. It also helped them address user needs and improve dictionaries. Other lexicographers, such as Atkins and Rundell (2008), denied the need for any lexicographic theory and highlighted the importance of some linguistic theories only as aiding tools for lexicographers.

They particularly referred to lexical semantics, cognitive theory, corpus linguistics and pragmatics as helping tools for lexicographers. It is worth mentioning that the history of lexicography precedes the history of lexical semantics, which used to rely on dictionaries to derive information about polysemy and semantic change (Geeraerts, 2013a).

The results of the three experiments conducted in this dissertation show the importance, but the insufficiency, of relying on theoretical information while building a dictionary entry with soundly delineated senses and representative examples. Cognitive semantic theories offer solutions to several lexicographic problems. Frame Semantics, for instance, has been used to inform decisions on sense separation and inclusion of information in dictionary entries based on a three-layer valence description of words (Atkins and Bouilon, 2006; Atkins, Rundell and Sato, 2003). Similarly, Image Schema contributed to the separation of the various senses of a preposition based on spatial and non-spatial taxonomy of the preposition's use (Xu and Lou, 2015). Prototype Theory is also effective in separating the prototypical core senses of a word from peripheral ones and explaining how peripheral meanings are derived from core ones (Jiang and Chen, 2015).

Nevertheless, the application of cognitive linguistic theories in lexicography has drawbacks, despite its advantages. First, cognitive linguistics theories rely heavily on introspective analysis. Intuitive judgment and introspection do not suit the supposedly unbiased linguistic judgment and evaluation (Dąbrowska, 2016). It can be said that Frame Semantics attempts to address this challenge by relying on the lexico-syntactic environment of the target words to separate senses and group lexical units within the same frame. Second, considerable methodological challenges are imposed by the different methods followed by each cognitive linguistic theory and the absence of a comprehensive theory or an approach in lexicography to contain the research work (Abdelzaher, 2021; Bogaards, 2013).

Ostermann (2015) proposed a framework for the systematic application of cognitive linguistic theories in order to improve lexicographic practice. The first step before conducting the analysis is the selection of any aspect of traditional lexicography. The choice of a compatible theory that can help lexicographers take better decisions follows. For instance, Prototype Theory has been fruitful in representing polysemy, whereas Conceptual Metaphor Theory has been effectively employed to define abstract concepts. The second step is deciding on the linguistic data to which the chosen theory should be applied and the lexicographic aspect explored. Linguistic data

includes the exploration of a specific semantic field, lexical unit, single words or idioms. The third step is the practical use of the major tenets of the chosen theory in the analysis of the data. The result of using a cognitive semantic theory in lexicographic practice is usually an edited dictionary entry in which a new feature is added or an old feature is improved. Therefore, the new entry should be evaluated by comparing it to existing dictionary entries in order to assess the improvement qualitatively.

Several lexicographic studies depend only on qualitative comparison between their theory-based modification of an entry and the original dictionary entry to evaluate the improvement, such as Abdelzaher and Tóth (2020), Dalpanagioti (2018, 2019), Fuertes-Olivera and Velasco-Sacristán (2012), Halas (2016) and Molina (2008). This widens the gap between the perspective of theoretical linguistics and that of field lexicographers on improving dictionaries. For lexicographers, decisions on what information to include in a dictionary should be user-oriented. Decisions should be informed either by studying the profile of the target user or by conducting user-based research on dictionary use (Atkins and Rundell, 2008).

Ostermann's framework (2015) addressed this problem by adding another evaluative step that targets dictionary users. This step involves exposing a target group of language learners to the edited entries and another control group to traditional entries. Then, both groups should perform the same task so that their performance can be compared. This step relatively bridges the gap between the theory-based improvements suggested by linguists and the user-oriented decisions taken by field lexicographers. In practical terms, it tests the influence of theory-based modifications on the performance of ESL learners. This last step is consistent with Cowie's (2007) enumeration of the various aims of user-based lexicographic research. Previous research used to focus on the perceived importance of specific linguistic information by dictionary users (addressed in Experiments I, II and III in this study), the expectations of users regarding a dictionary, the performance of a dictionary-related task (addressed in Experiments I, II and III in the current study) such as translation, the evaluation of a dictionary feature in comparison with another and the exploration of the suitability of different dictionary types to the learning process (addressed in Experiment I in the present study).

Although the proposed hybrid entry in the current study combines theoretical and lexicographic merits, several limitations impede the creation of similar entries for MLDs. First, the limited coverage of the FN database is a significant challenge to the use of FN as a basis for an MLD.

Also, at the macrostructural level, FN's wordlist does not follow an alphabetical or frequency-based detection of words. The frame-based recording of lexical units makes it impossible to find a relatively complete wordlist for a single letter or topic. Moreover, several words of the most frequent ones are missing from the database. Complicating matters, several senses of polysemous words may be missing because their frames have not been created yet. This limits the effectiveness of the sense delineation method, i.e., a word sense can be delineated in corpus examples but not added to the database before the creation of a suitable frame. In addition, the third experiment unveiled the limited value of the FN examples for the participants, especially if compared to OALD's.

The second limitation that impedes the creation of an MLD based on FN is the replacement of frame names with guidewords. Adopting guidewords from other dictionaries or a simplified version of the names is challenging, as discussed in Experiment II. Frame names were not composed for educational purposes. Therefore, they may be relatively long to describe a scenario, complicated and use C1 words to describe A2 lexical units or use unfamiliar word combinations. This is partly due to the fact that some of the very psychologically basic concepts are very challenging to explain (i.e., archetypes). Some frame names require prior knowledge about cognitive semantic theories (e.g., BOUNDED_REGION, PROXIMITY_IMAGE_SCHEMA), causative-inchoative alternations (e.g., PROCESS_END and CAUSE_TO_END, MAKE_NOISE and CAUSE_TO_MAKE_NOISE) or cognitive psychology (e.g., EMOTIONS_BY_STIMULUS, EXPERIENCER_FOCUSED_EMOTION, STIMULUS_FOCUS). Also, frame names do not usually have the same priming effect as guidewords (especially when the guideword is collocated with the target word). In addition, guidewords are usually specific to the word, but frame names can be shared between antonyms, co-hyponyms or LUs that are not related through conventional semantic relations.

Therefore, improving the FN database for lexicographic use may require integrating knowledge from other cognitive approaches and employing recent advances in lexicography. Crowdsourcing information has been successfully implemented in institutional MLDs such as *Collins Cobuild* and *Macmillian Open Dictionary.* The collaborative-institutional type would allow the FN team to consider the user's needs without affecting the quality of the information in the database. Checking the suggestions of the users regarding the simplification of frame names or the addition of a frame or an LU may change the perspective of the FN team and enhance the usability of the

database by language users and teachers. Also, considering the GDEX score of the provided examples may widen the scope of using the annotated examples in FN.

The structure of the WN database is more familiar to linguists and lexicographers, more compatible with ontological resources and simpler than that of the FN database. WN nests senses in conventional POS nets and uses lexical semantic relations to organize the database. This should enhance the usability of the database even by (advanced) learners. However, WN is challenged by the so-called 'tennis problem'. There is no formal semantic relation that can relate 'racquet', 'ball' and 'net' to each other (Fellbaum, 1998). The FN database, however, solves this problem by grouping words in the schematic representation they express (i.e., frames). This has been one of the reasons that led scholars to integrate FN and WN information in the same resource or to perform the same task. For instance, Baker and Fellbaum (2006) employed WN's paradigmatic relations and FN's syntagmatic relations to improve text understanding. The WN-based and FN-based information enriched the semantic annotation of the text with lexical relations and valence description. They linked LUs in the text to their frames and synsets.

In this regard, several attempts have been made to increase the coverage of the FN database or to create similar but richer databases. For instance, Bryl et al. (2012) were motivated by enriching the lexical coverage of the FN database. They mapped the lexical fillers of FEs to WN synsets and added this information to their proposed frame-based repository of senses. Also, Abdelzaher (2017) mapped violence-related LUs in FN to their corresponding synsets in WN to enrich the coverage of FN's violence-expressing LUs. The study also used a corpus to validate the use of the synonymous word as evocative of the same frame triggered by the FN target word. WN allowed increasing the coverage of the proposed lexicon with 34 LUs (based on WN) in addition to the 150 LUs retrieved from FN.

The new types of dictionaries in lexicography offer two solutions for integrating information from different dictionaries, i.e., portals and aggregators. Portals may not be the ideal solution because referring the user to the FN database will impose the previously discussed challenges. Aggregators would facilitate the chance of selective choice of relevant information from different lexicographic resources. Including lexicographic information from FN, WN, and MLDs in an aggregator can be beneficial to (advanced) language users.

Given the availability of multilingual framenets, bilingual frame-based dictionary entries can be proposed and exploited, whether in lexicographic classrooms or for translation purposes. Jódar-

Sánchez (2019) discussed the possibility of using the lexical entries of the Spanish framenet in teaching vocabulary to L2 learners and the use of the valence description in teaching grammar. In addition, the annotated examples can serve as ideal templates for correct sentence production. The inclusion of frame information in the definitions is also expected to meet the comprehension needs of ESL learners better than traditional dictionaries (Ostermann, 2015).

Moreover, Hasegawa et al. (2011) explained the role of frame-to-frame relations in helping users perform paraphrasing tasks effectively. They provided qualitative examples of the use of co-lexical units, frame-to-frame relations and grammatical information in FN in effective and systematic paraphrasing. To elaborate, co-lexical units can be used interchangeably to convey similar meanings. *Want*.v and *eager*.a, for instance, activate the DESIRING frame and occur with the same frame elements. They can be used interchangeably to paraphrase sentences using the "intersubstitutability of synonymous expressions". Lexical units belonging to frames related by an FN relation are good candidates for paraphrasing tasks, too. For instance, the Inheritance relation between INHIBIT_MOVEMENT and IMPRISONMENT allows the conceptual structure of the parent frame to be inherited by the child frame with some additional specifications. Moreover, available grammatical information in FN, such as the different valence patterns, is beneficial to paraphrasing tasks. The valence patterns of *rude*.a in the frame SOCIAL_INTERACTION_EVALUATION display several ways in which "Agent" and "Evaluee" are realized.

Although these suggestions are promising and reasonable, the current structure of the database remains a challenge to using it for lexicographic purposes. Rundell (2008) referred to the orientation to user needs as the main reason for Hornby's innovations in learners' dictionaries (e.g., vocabulary control, simple definitions, grammatical information, pedagogical examples and phraseology inclusion). Any improvement in MLDs should be measured in terms of the description and representation of language use. FN uses the BNC and U.S. newswire texts provided by the Linguistic Data Consortium to guarantee the authenticity of the description of the use of each lexical unit. Although the description of language use is consistent with Hornby's innovations, the presentation of this information in FN presupposes advanced knowledge and sophisticated navigation skills.

Looking up any word in FN retrieves a table of hyperlinks which include the label of the frames evoked by the lexical unit, the status of the lexical unit, a lexical entry report and an annotation

report for each lexical unit. The retrieval of dozens of hyperlinks in the same window after looking up a word can be very confusing for students who attempt to find only the meaning of a word. Complicating matters, none of the retrieved information includes the sense of the looked-up word. In addition, users need either to click on each lexical entry report to find the different senses of the target word or to select specific entries based on the displayed frame label. The information about a word is scattered across different LU entries, annotated examples, valency tables and frames entries. Users need to keep navigating the various sections in each entry to complete the lexicographic description of a word.

For instance, in the valency description section in the LU entry, users can view the annotated sentences which realize a specific frame element or a specific valence pattern. However, the valency description does not include any definition of the frame elements. The user needs to navigate through the frame entry to find the definition of the frame element, and then move back to the lexical entry report and check how the frame element is realized. Prior knowledge about the abbreviations (e.g., DNI, CNI) should be acquired from the FN glossary too. This complicates the mediostructure of the database. The findability of information in FN will be a dilemma for students who lack theoretical knowledge about FS, semantic roles and valency.

Accordingly, the use of any framenet database in its current form needs at least some modifications at the megastructure. Instead of altering the microstructure of the database, which needs massive computational effort and time and will negatively impact any relevant NLP application, the FN teams can make use of the traditional front matters section in dictionaries. In the electronic web-based form, the "about" section can serve this purpose. It can include information about the access structure of the database. This section can inform users about the two access structures of the database through the LU index and the frame index, explains how information is presented at the microstructure level and lists the limitations of the macrostructure of the database. This should increase the usability of the database by ESL learners in lexicographic contexts.

However, changing the content or the interface of a lexicographic resource is not enough to make it usable by the target group. The target users of a dictionary or a reference book should have the necessary skills to use and understand the information in a lexicographic reference. Therefore, professional lexicographic training at schools and higher education institutions is a necessity Tóth, Márkus and Pődör (2022) explored how lexicography and dictionary didactics are taught in

the university education of English teachers in Hungary. They argued that lack of knowledge about quality dictionaries at the beginning of the learning process results in learners' inability to select the right sources of information from the vast number of websites and search results on the Internet. They examined the practice of some Hungarian higher education institutions in teaching and training lexicography. They reported that a specialised course is occasionally organised for students, but lexicography and dictionary didactics are not taught in a more focused form. Students have a low appreciation of dictionaries, consider print dictionaries obsolete, and use free online dictionaries. The most used dictionaries reported in that study were a bilingual dictionary (Magay-Országh) and an MLD (OALD).

In the current study, we received a total of 150 responses in the three experiments. All participants were affiliated with the English Department at the Institute of English and American Studies, Debrecen University. Students' perplexity in the first and second experiments and their uncertainties regarding the applicability of most of the senses on the test sentences in the last experiment doubt their effective general use of dictionaries. Also, their preference to use *Google Translate* for lexicographic purposes, familiarity with collaborative dictionaries such as the *Urban Dictionary* and reluctant use of several institutional MLDs reveal their lack of lexicographic awareness (especially in the 1st and 2nd years). According to Dringó-Horváth and Márkus (2022), graduates from language departments at Hungarian Higher Education Institutes showed similar preferences for using translation software and search engines for dictionary purposes. Moreover, the participants did not acknowledge the importance of teaching dictionary skills and referred to acquiring dictionary skills through self-learning.

Teaching lexicography, especially at the university level, should simplify the differences between the complexity of meaning in the human mind and the need to define words for language learners. Users will be aware of the fact that dictionaries attempt to give them access to the potential meanings and uses of a word. Therefore, dictionary users would not have unfulfilled expectations while consulting dictionaries. Teaching what a dictionary is, how it should be consulted and, equally important, the complexity and richness of word meaning would help learners make the best use of a dictionary and decrease their perplexity. It will also prepare the students to accept the inapplicability of having strict one-to-one correspondences between dictionary senses and word uses. Therefore, integrating cognitive-linguistic information in teaching lexicography is recommended to help students make the best of dictionaries.

Dictionaries

Hornby, A.S., Gatenby, E.V. and Wakefield, H. (1942). *Idiomatic and Syntactic English Dictionary*. Kaitakusha, Tokyo.

Hornby, A.S., Gatenby, E.V. and Wakefield, H. (1948). *A Learner's Dictionary of Current English*. Oxford University Press

Hornby, A.S., Gatenby, E.V. and Wakefield, H. (1952). *The Advanced Learner's Dictionary of Current English*. Oxford University Press.

Hornby, A.S. (1954). *A Guide to Patterns and Usage in English*. Oxford University Press.

Hornby, A. S., Cowie, A. and Lewis, J. (1974). *Oxford Advanced Learner's Dictionary of Current English*. (third edition). Oxford University Press.

Palmer, H.E. (1938). *A Grammar of English Words*. Longmans, Green.

Proctor, B. (1978). *Longman Dictionary of Contemporary English*. Longmans, Green.

Proctor, B. (1995). *Cambridge International Dictionary of English*. Cambridge University Press

Rundell, M. (2002) *Macmillan English Dictionary for Advanced Learners*. Macmillan Education.

Sinclair, J. (1987). *Collins COBUILD English Language Dictionary*. HarperCollins publishers.

West, M.P. and Endicott, J.G. (1935). *The New Method English Dictionary*. Longmans, Green.

References

Abdelzaher, E. (2017). Compiling a cognition-based thematic monolingual dictionary. *Cognitive Linguistic Studies*, *4*(2). doi 10.1075/cogls.00007.abd

Abdelzaher, E. M. (2021). Cognitive Linguistics and Digital Lexicography. In W. Xu and J. Taylor (eds). *The Routledge Handbook of Cognitive Linguistics*. (pp. 568–584). https://doi.org/10.4324/9781351034708-38

Abdelzaher, E. and Tóth, Á. (2020). Defining Crime: A multifaceted approach based on Lexicographic Relevance and Distributional Semantics. *ARGUMENTUM*, 16, 44–63. https://doi.org/10.34103/argumentum/2020/4

Abecassis, M. (2007). Is lexicography making progress? On dictionary use and language learners' needs. *Lexikos*, 17. https://doi.org/10.5788/17-0-555

Ackerman, F., Kay, P., and O'Connor, M. C. 2014. Charles J. Fillmore. *Language*, 90: 755-761.

Adamska-Sałaciak, A. (2013). Issues in compiling bilingual dictionaries. In Jackson, H. (Ed.) *The Bloomsbury companion to lexicography*.

Adamska-Sałaciak, A. (2019). Lexicography and Theory: Clearing the Ground. *International Journal of Lexicography*, *32*(1), 1–19. https://doi.org/10.1093/ijl/ecy017

Agirre, E. and Edmonds, P. (2007). *Word sense disambiguation: Algorithms and applications*. Springer Science and Business Media.

Alekse, L. (2017). *Distinguishing Between Polysemy and Homonymy: A Critique of a Common Dictionary Approach*. (Thesis)

Andor, J. 2010. Discussing Frame Semantics: The State of the Art. An Interview with Charles Fillmore. *Review of Cognitive Linguistics*, 8: 157-176.

Apresjan, J. D. (1974). Regular polysemy. *Linguistics*, 12(142). doi:10.1515/ling.1974.12.142.5

Atkins, S. and Bouillon, P. (2006): Relevance in dictionary making: Sense indicators in the bilingual entry. In: Bowker, L. (ed.): *Lexicography, terminology, and translation: Textbased studies in honour of Ingrid Meyer*. Ottawa: University of Ottawa Press, 25–43, https://doi.org/10.2307/j.ctt1ckpgs3.6.

Atkins, S. (2008). Theoretical Lexicography and its Relation to Dictionary-making. In Fontenelle, T. (Ed). *Practical Lexicography: A Reader*. Oxford: Oxford University Press.

Atkins, S. and Rundell, M. (2008). *The Oxford guide to practical lexicography*. Oxford University Press.

Atkins, S., Rundell, M. and Sato, H. (2003): The contribution of FrameNet to practical lexicography. *International Journal of Lexicography,* 16(3), 333–357, https://doi.org/10.1093/ijl/16.3.333.

Baker, C. (2012), Framenet, current collaborations and future goals. *Language Resources and Evaluation, 46*(2), pp. 269–286.

Baker, C., Fillmore, C. and Lowe, J. 1998. The Berkeley FrameNet Project. *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, 1, 86-90.

Bergenholtz, H. and Agerbo, H. (2018). A typology of lexicographical tools based on information needs and user types. *Lexicography*, 5(2), 97–121. https://doi.org/10.1007/s40607-018-0050-1

Béjoint, H. (1994). *Tradition and innovation in modern English dictionari*es. Oxford: Oxford University Press.

Bergenholtz, H. and Gouws, R. H. (2017). Polyseme selection, lemma selection and article selection. *Lexikos*, *27*, 107-131.

Boas, H. and Dux, R. 2017. From the past into the present: From case frames to semantic frames. *Linguistics Vanguard*, 3: 1-14

Bogaards, P. (2013). A history of research in lexicography. In Jackson I. (Ed.), *The Bloomsbury Companion to Lexicography* (pp. 19–31).

Britton, B. K. (1978). Lexical ambiguity of words used in English text. *Behavior research methods and Instrumentation*, 10(1), 1-7.

Brocher, A., Foraker, S. and Koenig, J.-P. (2016). Processing of Irregular Polysemes in Sentence Reading. Journal of Experimental Psychology: *Learning, Memory, and Cognition*.

Buitelaar, P. (1998). *CORELEX: Systematic polysemy and underspecification*. (Dissertation)

Buitelaar, P. (2010). Ontology-based semantic lexicons: Mapping between terms and object descriptions. *In Ontology and the lexicon: A natural language processing perspective* (pp. 212-223). Cambridge: Cambridge University Press.

Buyko, E., Beisswanger, E. and Hahn, U. (2010). The GENEREG corpus for gene expression regulation events - an overview of the corpus and its in-domain and out-of-domain interoperability. *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, 2662-2666.

Carston, R. (2021). Polysemy: Pragmatics and sense conventions. *Mind and Language, 36*(1), 108-133.

Chan, A. (2014). How can ESL students make the best use of learners' dictionaries?. *English Today*, 30, pp 33-37 doi:10.1017/S0266078414000248

Chi, A. (2013). Researching Pedagogical Lexicography. In Jackson, H. (Ed). *The Bloomsbury companion to lexicography*.

Cinková, S. and Hanks, P. (2010). Validation of Corpus Pattern Analysis–Assigning pattern numbers to random verb samples.

Chow, I.C. and T. Wong (2006), Axiomatizing relational network for knowledge engineering - exploring wordnet and framenet. *Proceedings of the 2006 IEEE International Conference on Information Reuse and Integration*, Hawaii, USA.

Considine, J. (1997). Special Feature: Discussion 2 Etymology and the Oxford English Dictionary: a response. *International Journal of Lexicography, 10*(3), 234-236.

Cowie, A. P. (2007). *English dictionaries for foreign learners: A history*. Oxford: Oxford University Press.

Cruse, A. (2000). *Meaning in Language. An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.

Cruse, A. (2002). Hyponymy and its variants. In R. Green and C. Bean (eds). *The semantics of relationships: An interdisciplinary perspective.* Springer

Cruse, A. (2003). The lexicon. In M. Aronoff and J. Ress-Miller (eds). *The Handbook of Linguistics* (238-264).

Cruse, D. A. (1986). *Lexical semantics*. Cambridge, Great Britain: Cambridge University Press.

Cruse, D. A. (2017). The lexicon. *In M. Aronoff and J. Ress-Miller (eds). The Handbook of Linguistics.* 235-254.

Dąbrowska, E. (2016). Cognitive Linguistics' seven deadly sins. *Cognitive linguistics*, 27(4),

479-491.

Dalpanagioti, T. (2018). A frame-semantic approach to co-occurrence patterns: A lexicographic study of English and Greek motion verbs. *International Journal of Lexicography*, 31(4), 420-451.

Dalpanagioti, T. (2019). From corpus usages to cognitively informed dictionary senses: reconstructing an MLD entry for the verb float. *Lexicography*, 6(2), 75-104.

De Schryver, G. M., Chishman, R. and da Silva, B. (2019). An overview of Digital Lexicography and directions for its future: An interview with Gilles-Maurice de Schryver. *Calidoscópio*, 17(3), 659-683.

De Schryver, G. M. and Nabirye, M. (2018a). Corpus-driven Bantu Lexicography Part 2: Lemmatisation and rulers for Lusoga. *Lexikos*, 28, 79–111. https://doi.org/10.5788/28-1-1458

Dobrić, N. (2015). Three-factor prototypicality evaluation and the verb look. *Language Sciences*, 50, 1-11.

Dziemianko, A. (2015). An insight into the visual presentation of signposts in English learners' dictionaries online. *International Journal of Lexicography*, *29*(4), 490-524.

Dziemianko, A. (2016). An insight into the visual presentation of signposts in English learners' dictionaries online. *International Journal of Lexicography*, *29*(4), 490-524.

Dziemianko, A. (2017). Dictionary entries and bathtubs: Does it make sense?. *International Journal of Lexicography*, *30*(3), 263-284.

Dziemianko, A. (2018). Electronic dictionaries. In P. Feurtes-Olivera (Ed). *The Routledge Handbook of Lexicography*.

Faaß, G. (2018). Lexicography and corpus linguistics. In P. Feurtes-Olivera (Ed). The Routledge Handbook of Lexicography (pp. 123–137). https://doi.org/10.4324/9781315104942-9

Fellbaum, C., D. Gross and K. Miller (1993) Adjectives in WordNet. In Miller (Ed). *Five Papers in WordNet.* Princeton University, pp. 26–39.

Fillmore, C. (1968a). The Case for Case. *Form and Meaning in Language: Volume 1 Papers on Semantic Roles.* 21-119.

Fillmore, C. (1968b). Lexical Entries for Verbs. *Foundations of Language*, 4: 373-393.

Fillmore, C. (1969). Verbs of judging: An exercise in semantic description. *Paper in Linguistics*, 1: 91-117.

Fillmore, C. (1970). Subjects, Speakers and Roles. *Synthese* 21, 251-274.

Fillmore, C. (1972). A grammarian looks to sociolinguistics. *Report of the Twenty-Third Annual Round Table Meeting on Linguistics and Language Studies*, 273-287.

Fillmore, C. (1975). An Alternative to Checklist Theories of Meaning. *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, 123-131.

Fillmore, C. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech,* 280: 20-32.

Fillmore, C. (1977a). The case for case reopened. In P. Cole and J. Sadock (eds.). *Grammatical relations.* 59-81.

Fillmore, C. (1977b). Scenes-and-frames Semantics. ". In A. Zampolli, (ed.). *Linguistic Structures Processing.* 55-81.

Fillmore, C. (1985). Frames and the semantics of understanding. *Quaderni di semantica*, 6: 222-254.

Fillmore, C. (1988). The Mechanisms of "Construction Grammar". *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, 35-55

Fillmore, C. (1992). "Corpus linguistics" or "Computer-aided armchair linguistics". In J. Svartvik (ed.). *Directions in Corpus Linguistics*, 35-60.

Fillmore, C. (2003). Double-decker definitions: The role of frames in meaning explanations. *Sign Language Studies*, 3: 263-295.

Fillmore, C. and Atkins, S. (1992). Towards a frame-based organization of the lexicon: the semantics of RISK and its neighbors. In A. Lehrer and E. Kittay (eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization*, 75-102.

Fillmore, C. J., Johnson, C. R., and Petruck, M. R. L. (2003). Background to FrameNet. *International Journal of Lexicography*, 16: 235-250.

Fontenelle, T. (Ed.). (2008). *Practical lexicography: a reader*. Oxford: Oxford University Press.

Fraser, B. L. (2008). Beyond Definition: Organising semantic information in bilingual dictionaries. *International Journal of Lexicography, 21*(1), 69-93.

Fuertes-Olivera, P. A. and Velasco-Sacristan, M. (2012). Translation metaphor in business/economics dictionary articles: what the theory says and what the lexicographers should do. *Metaphors and mills*. 155-173.

Geeraerts, D. (2001). The definitional practice of dictionaries and the cognitive semantic conception of polysemy. *Lexicographica*, 17, 6– 21.

Geeraerts, D. (2013a). Lexical semantics from speculative etymology to structuralist semantics. In *The Oxford handbook of the history of linguistics*.

Giora, R. (2003). *On our mind: Salience, context, and figurative language*. Oxford: Oxford University Press.

Gouws, R. H. (2018). Dictionaries and access. In P. Feurtes-Olivera (Ed). *The Routledge Handbook of Lexicography* (pp. 43–58). https://doi.org/10.4324/9781315104942-4

Gouws, R. H. and Tarp, S. (2017). Information overload and data overload in

lexicography. *International Journal of Lexicography, 30*(4), 389-415.

Grochocka, M. (2008). The usefulness of the definitions of abstract nouns in OALD7 and NODE. *Poznań Studies in Contemporary Linguistics*, *44*(4), 469– 501.

Halas, A. (2016). The application of the prototype theory in lexicographic practice: a proposal of a model for lexicographic treatment of polysemy. *Lexikos*, 26(1), 124-144.

Hanks, P. (2004). Corpus pattern analysis. *Euralex Proceedings*. Lorient: Université de Bretagne-Sud.

Hanks, P. (2013). Lexicography from Earliest Times to the Present. In *The Oxford Handbook of the History of Linguistics* (pp. 503–536).

Hanks, P. (2019). Lexicography. In R. Mitkov (ed). *The Oxford Handbook of Computational Linguistics* (2nd ed., Vol. 9780199276). https://doi.org/10.1093/oxfordhb/9780199276349.013.0003

Hasegawa, Y., Lee-Goldman, R., Kong, A., and Kimi, A. (2011). FrameNet as a resource for paraphrase research. *Constructions and Frames*, 3: 104-127.

Hovy, E., Marcus, M., Palmer, M., Ramshaw, L. and Weischedel, R. (2006, June). OntoNotes: the 90% solution. *Proceedings of the human language technology conference of the NAAL*.

Humbley (2018) Specialized dictionaries. In Fuertes-Olivera, P. A. (Ed.). *The Routledge handbook of lexicography.*

Hung, B. P. (2019). A cognitive linguistic approach to teaching English idioms to EFL students: Experimental results. *3L: Language, Linguistics, Literature*, *25*(2), 113– 126.

Ide, N. and Wilks, Y. (2007). Making sense about sense. In *Word sense disambiguation* (pp. 47-73). Springer, Dordrecht.

Jackson, H. (2003). *Lexicography: An Introduction*. Taylor and Francis Routledge

Jódar-Sánchez, J. A. (2019). FrameNet as a Resource to Teach Spanish as a Foreign Language. In *Teaching Language and Teaching Literature in Virtual Environments* (pp. 121-149). Springer, Singapore.

Jiang, G. and Chen, Q. (2017). A micro exploration into learner's dictionaries: A prototype theoretical perspective. *International Journal of Lexicography*, 30(1), 108-139.

Kilgarriff, A. (1992). Dictionary word sense distinctions: An enquiry into their nature. *Computers and the Humanities*, *26*(5), 365-387.

Kilgarriff, A. (1997). I don't believe in word senses. *Computers and the Humanities*, *31*(2), 91-113.

Kilgarriff, A. (1998). The hard parts of lexicography. *International Journal of Lexicography*, 11(1), 51–54. https://doi.org/10.1093/ijl/11.1.51

Kilgarriff, A. (2005). Putting the corpus into the dictionary. In Proceedings *Meaning Workshop*.

Kilgarriff, A. (2007). Googleology is Bad Science. *Computational Linguistics, 3 3*(1), 147-151.

Kilgarriff, A. (2007). Word sense disambiguation. In P. Agirre, E. and Edmonds (Ed.), *Word sense disambiguation: Algorithms and applications (Vol. 33).* Springer Science and Business Media.

Kilgarriff, A. (2013). Using corpora as data sources for dictionaries. In H. Jackson (ed*). The Bloomsbury Companion to Lexicography*. London: Bloomsbury, 77-96.

Kilgarriff, A. and Kosem, I. (2012). Corpus tools for lexicographers. In S. Granger and M. Paquot (eds). *Electronic Lexicography*.

Kilgarriff, A. and Rundell, M. (2002). Lexical profiling software and its lexicographic applications-a case study. *In Proceedings of EURALEX* (pp. 807-818).

Kilgarriff, A., Husák, M., McAdam, K., Rundell, M. and Rychlý, P. (2008). GDEX: Automatically finding good dictionary examples in a corpus. In *Proceedings of the XIII EURALEX international congress* (pp. 425-432). Barcelona, Spain: Documenta Universitaria.

Klein, D. E. and Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language*, 45, 259-282. doi:10.1006/jmla2001.2779

Kövecses, Z. (1986). *Metaphors of anger, pride and love: A lexical approach to the structure of concepts*. Amsterdam: John Benjamins.

Kövecses, Z. (1991). Happiness: A definitional effort. *Metaphor and Symbol*, *6*(1), 29– 47. DOI: 10.1207/s15327868ms0601_ 2

Krishnamurthy, R. (1996). *The process of compilation*. In J. Sinclair (ed). Looking up: An account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary. London: HarperCollins Publisher.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.

Lakoff, G. 2014. Charles Fillmore, discoverer of frame semantics, dies in SF at 84: he figured out how framing works. *Review of Cognitive Linguistics* 12: 251-257

Lakoff, G. and Johnson, M. (1980). Conceptual metaphor in everyday language. *The journal of Philosophy, 77*(8), 453-486.

Landau, S. (1993). *Dictionaries: The art and craft of lexicography*. Oxford: Oxford University Press.

Landes, S., Leacock, C. and Tengi, R. I. (1998). Building semantic concordances. In G. Miller (ed). *WordNet: An electronic lexical database*, 199-216.

Langacker, R. W. (1999). *Grammar and conceptualization* (Vol. 14). Walter de Gruyter.

Langacker, R. (2017). *Ten lectures on the basics of cognitive grammar*. Brill.

Laparra, E., G. Rigau, and M. Cuadros (2010). Exploring the integration of wordnet and

framenet. *Proceedings of the 5th Global WordNet Conference*, Mumbai, India.

Lenci, A. (2010). The life cycle of knowledge. In Huang et al. (eds). *Ontology and the lexicon: A natural language processing perspective* (pp. 241-257). Cambridge University Press.

Leroyer, P. (2018). Dictionaries for text reception. In P. Feurtes- Olivera (Ed.), *The Routledge handbook of lexicography*, (pp. 250– 266). London and New York: Routledge.

Lew, R. (2013). Identifying, ordering and defining senses. In Jackson, I. (Ed). *The Bloomsbury companion to lexicography*, 284-302.

Lew, R. and Pajkowska, J. (2007). The effect of signposts on access speed and lookup task success in long and short entries. *Horizontes de Lingüística Aplicada* 6.2: 235-252.

Lewandowska-Tomaszczyk, B. (2007). *Polysemy, prototypes, and radial categories*. The Oxford handbook of cognitive linguistics, 139-169.

Litkowski, K. (2014). Pattern dictionary of English prepositions. *In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. 1, 1274-1283.

Lu, H. and Wei, X. (2019). Structuring polysemy in English learners' dictionaries: A prototype theory- based model. *International Journal of Lexicography*, 32(1), 20– 37.

Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.

Márkus, K. P. and Pődör, D. (2021). Lexikográfia és szótárdidaktika a Károli Gáspár Református Egyetem Angol Nyelvészeti Tanszékén. *Iskolakultúra*, 31(5), 92-107.

Martínez Alonso, H. (2013). *Annotation of regular polysemy: an empirical assessment of the underspecified sense* (Doctoral dissertation).

Mheta, G. (2017). A cognitive analysis of metaphor in Shona terminological dictionaries. *Southern African Linguistics and Applied Language Studies*, 35(3), 311-320.

Miller G., C. Leacock, T. Randee, R. Bunker. 1993. A semantic concordance. In *Proceedings of the 3rd DARPAWorkshop on Human Language Technology*, Plainsboro, NJ, 303–308.

Miller, G. A. (1998). *WordNet: An electronic lexical database*. MIT press.

Miller, G. A., Leacock, C., Tengi, R. and Bunker, R. T. (1993). A semantic concordance. In Human Language Technology: *Proceedings of a Workshop Held at Plainsboro*, New Jersey.

Molina, C. (2008). Historical dictionary definitions revisited from a prototype theoretical standpoint. *Annual Review of Cognitive Linguistics*, *6*(1), 1– 22.

Moon, R. (1996). *The analysis of meaning*. in J. Sinclair (ed). Looking up: An account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary. London: HarperCollins Publisher.

Murray, J. (1887). *A new English Dictionary: On historical principles*. Oxford: At the Clarendon Press.

Nelson, K. (2020). Informing Lexicographic Choices Through Corpus and Perceptual Data. *International Journal of Lexicography*, 33(3), 251-268.

Nesi, H. and Tan, K. H. (2011). The effect of menus and signposting on the speed and accuracy of sense selection. *International Journal of Lexicography*, *24*(1), 79-96.

Ooi, V. B. Y. (2021). Issues and prospects for incorporating English use in Japan into the dictionary. *Asian Englishes*, *23*(1), 62–78. https://doi.org/10.1080/13488678.2021.1876952

Ostermann, C. (2015). *Cognitive lexicography: A new approach to lexicography making use of cognitive semantics.* Walter de Gruyter GmbH and Co KG. doi: 10.1515/9783110424164

Palmer, NG and Tang (2007) Evaluation of word sense disambiguation systems. In Agirre and P. Edmonds (eds.), *Word Sense Disambiguation: Algorithms and Applications*, 29–46

Pedersen, T., Patwardhan, S. and Michelizzi, J. (2004, May). WordNet:: Similarity: measuring the relatedness of concepts. *Demonstration papers at HLT-NAACL 2004* (pp. 38-41).

Petten, C. (2006). Lexical ambiguity resolution. In L. Nadel (ed) *Encyclopedia of cognitive science*. Wiley.

Philpot, A., Fleischman, M. and Hovy, E. H. (2003). Semi-automatic construction of a general purpose ontology. In *Proceedings of the International Lisp Conference*. New York, NY. Invited.

Philpot, A., Hovy, E. and Pantel, P. (2005). The omega ontology. In *Proceedings of OntoLex 2005-Ontologies and Lexical Resources*.

Piaget, J. (1976). *Piaget's Theory*. In: Inhelder, B., Chipman, H.H., Zwingmann, C. (eds) Piaget and His School. Springer Study Edition. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-46323-5_2

Pragglejaz Group. (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol, 22*(1), 1– 39.

Prévot, L., Huang, C. R., Calzolari, N., Gangemi, A., Lenci, A. and Oltramari, A. (2010). Ontology and the lexicon: a multi-disciplinary perspective. In *Ontology and the lexicon: A natural language processing perspective* (pp. 3-24). Cambridge University Press.

Prinsloo, D. J. (2015). Corpus-based lexicography for lesser-resourced languages-maximizing the limited corpus. Lexikos, 25, 285–300. https://doi.org/10.5788/25-1-1300

Ptasznik, B. and Lew, R. (2014). Do menus provide added value to signposts in print monolingual dictionary entries? An application of linear mixed-effects modelling in dictionary user research. *International Journal of Lexicography, 27*(3), 241-258.

Ptasznik, B. and Lew, R. (2019). New-line and Run-on Guiding Devices in Print Monolingual Dictionaries for Learners of English. *Lexikos*, 29, 180-198.

Rosch, E. (1978). Principles of Categorization. In E. Rosch and B. Lioyd (eds). *Cognition and Categorization*. Lawrence Erlbaum associates publishers

Ruppenhofer, J., Ellsworth, M., Schwarzer-Petruck, M., Johnson, C. R. and Scheffczyk, J. (2016). *FrameNet II: Extended theory and practice*. International Computer Science Institute.

Sambre, P. 2010. Framing from grammar to application. *Belgian journal of linguistics*, 24: 1-15.

Smirnova, A. (2021). In Awe of God, Nature and Technology: A Lexical Approach to the Differentiation of Emotional Responses. *3L: Language, Linguistics, Literature*, *27*(4), 230–243. https://doi.org/10.17576/3L-2021-2704-16

Sinclair, J. (Ed.). (1987). *Looking up: An account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary*. London: HarperCollins Publisher.

Siqueira, M., de Oliveira, A. F. S., Hubert, D. D., de Almeida, G. F. and Brangel, L. M. (2009). Metaphor identification in a terminological dictionary. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos*, 17, 157– 174.

Spence, J. (2021). A Corpus Too Small: Uses of Text Data in a Hupa-English Bilingual Dictionary. *International Journal of Lexicography*, *34*(4), 413–436. https://doi.org/10.1093/ijl/ecab006

Speranza, M. and Magnini, B. (2010). Merging global and specialized linguistic ontologies. In *Ontology and the lexicon: A natural language processing perspective* (pp. 3-24). Cambridge University Press.

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24-31.

Sullivan, K. 2016. Integrating constructional semantics and conceptual metaphor. *Constructions and Frames*, 8:141-165.

Sweep, J. 2010. A frame-semantic approach to logical metonymy. *Constructions and Frames.* 2:1-32.

Tarp, S. (2018a). Lexicography as an independent science. In P. Feurtes-Olivera (Ed). *Routledge Handbook of lexicography*.

Tarp, S. (2018b). The concept of dictionary. In P. Feurtes-Olivera (Ed). *Routledge Handbook of lexicography.*

Tóth, A. (2008). *Perspectives on the lexicon*. Budapest: Akadémiai Kiado.

Tóth, Á. (2019). *Az Oxford 3000 definiáló alapszókincs és annak 2019. évi változása*. In: Fóris, Ágota; Bölcskei, Andrea; M., Pintér Tibor; Szoták, Szilvia; Tamás, Dóra Mária (szerk.) Nyelv, kultúra, identitás. Alkalmazott nyelvészeti kutatások a 21. századi információs térben : I. Terminológia, lexikográfia, fordítás. Budapest: Akadémiai Kiadó.

Tóth, Á., Márkus, K. and Pődör, D. (2022). *Lexikográfia és szótárdidaktika az angoltanárképzésben – helyzetkép*. In: Előry Vilma; Tóth Ágoston (eds.) A szótár az oktatásban:

A lexikográfiától a szótárhasználatig. Budapest: Tinta Könyvkiadó.

Tonelli, S. and D. Pighin (2009), New features for framenet: Wordnet mapping. *Proceedings of the thirteenth Conference on Computational Natural Language Learning*, Boulder, Colorado

Van der Meer, G. (1999). Metaphors and dictionaries: The morass of meaning, or how to get two ideas for one. *International Journal of Lexicography*, 12(3), 195– 208.

Wiegand, H. E., Beer, S. and Gouws, R. H. (2014). 3. Textual structures in printed dictionaries: An overview. In *Dictionaries. An International Encyclopedia of Lexicography* (pp. 31–73). https://doi.org/10.1515/9783110238136.31

Wild, K., Kilgarriff, A. and Tugwell, D. (2013). The Oxford Children's Corpus: Using a Children's Corpus in Lexicography. *International Journal of Lexicography*, 26(2). https://doi.org/10.1093/ijl/ecs017

Wójtowicz, B. (2016). Learner features in a new corpus-based Swahili dictionary. *Lexikos*, *26*, 402–415. https://doi.org/10.5788/26-1-1343

Xu, H. and Lou, Y. (2015). Treatment of the preposition to in English learners' dictionaries: A cognitive approach. *International Journal of Lexicography*, 28(2), 207-231.

Xia, L., Xia, Y., Zhang, Y. and Nesi, H. (2016). The corpora of China English: Implications for an EFL dictionary for Chinese Learners of English. *Lexikos*, *26*, 416–435. https://doi.org/10.5788/26-1-1342

Yasuda, S. (2010). Learning phrasal verbs through conceptual metaphors: A case of Japanese EFL learners. *Tesol Quarterly*, 44(2), 250-273.

Yurchenko, A., Lopukhina, A. and Dragoy, O. (2020). Metaphor Is Between Metonymy and Homonymy: Evidence From Event-Related Potentials. *Frontiers in Psychology*, 11, 2113.

Abstract

This dissertation addresses the lexicographic challenge of sense delineation. It explores the possible contributions of cognitive linguistics and linguistic ontologies to enhancing sense delineation and, accordingly, the presentation of senses in monolingual learners' dictionaries. The first part of this dissertation presents theoretical discussions of the recent advances in lexicography (e.g., corpus tools, user-generated content, web-based dictionary models), the presentation of meaning in linguistic ontologies (e.g., WordNet, CoreLex) and cognitive linguistic views on meaning (e.g., Charles Fillmore's Frame Semantics, Ronald Langacker's Cognitive Grammar, George Lakoff's Conceptual Metaphor Theory). The dissertation's main argument is that sense delineation in WordNet and the theoretical advances in cognitive linguistics can improve the process of sense delineation. In addition, the recent advances in lexicography can enhance the accessibility of the WordNet and FrameNet databases.

Three experiments were conducted to test the theoretical arguments made in the first chapters. The experiments assessed the influence of cognitive-linguistic, ontological and lexicographic approaches to sense delineation on university students' decoding and encoding performances in a dictionary consultation context. Whereas the first experiment tests only the senses delineated by FrameNet, WordNet and Oxford Learners Dictionary, the second experiment proposes a hybrid entry that uses the senses delineated by FrameNet, the signposts presented by Oxford Learners Dictionary and the hypernyms and hyponyms listed in WordNet. The third experiment focuses on the role of examples in clarifying word senses in lexicographic resources and the perceived applicability of the senses in an entry to specific example sentences.

A total of 150 students at the Institute of English and American Studies, Debrecen University, participated in the three experiments. Comparing the students' performances in the three groups in the first experiment showed the superiority of the FrameNet system to WordNet and Oxford Dictionary systems. Students in the FrameNet group showed the best encoding and decoding performances, the least perplexity levels and spent the shortest time on the consultation. The target group performed relatively better in the second experiment than the control group. The results advocate the effectiveness of integrating lexicographic information from different resources despite the complexity of the task. The integration of lexicographic data from FrameNet, WordNet and Oxford Learners Dictionary resulted in developing the most helpful dictionary entries for the participants despite the theoretical and practical challenges of creating

such entries. The third experiment, however, unveiled the drawbacks of the three resources while providing example sentences. It was evident that the one-to-one mapping between word senses and word uses is unattainable in various cases. Meaning extension, fuzzy meaning categories and diversity of profiles within the same conceptual base further complicate the process of delineating senses. This was reflected in the students' choice of multiple senses as the most applicable to each example sentence.

FrameNet lexicographers can increase the effective use of the database by using the new advances in lexicography. Crowdsourcing simpler frame names can enhance the accessibility of word senses in the database. Also, applying the Good Dictionary Example (GDEX) measure to the sentences before their inclusion in the database will help lexicographers exclude learner-challenging examples. As regards the WordNet database, the combined use of SemCor examples and GDEX may improve the presentation of the examples.

Keywords

FrameNet, Monolingual Learners' Dictionaries, Sense delineation, WordNet

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/88/2023.PL
Subject: PhD Publication List

Candidate: Esra Abdelzaher
Doctoral School: Doctoral School of Linguistics
MTMT ID: 10065286

List of publications related to the dissertation

Foreign language international book chapters (1)

1. **Abdelzaher, E.**: Cognitive Linguistics and Digital Lexicography.
   In: The Routledge Handbook of Cognitive Linguistics / edited by Xu Wen and John R. Taylor, Routledge, New York, 568-584, 2021, (Routledge handbooks in linguistics) ISBN: 9781138490710

Foreign language scientific articles in Hungarian journals (1)

2. **Abdelzaher, E.**, Tóth, Á.: Defining Crime: A multifaceted approach based on Lexicographic Relevance and Distributional Semantics.
   Argumentum (Debr.). 16, 44-63, 2020. EISSN: 1787-3606.
   DOI: http://dx.doi.org/10.34103/ARGUMENTUM/2020/4

Foreign language scientific articles in international journals (3)

3. **Abdelzaher, E.**: A classroom-based study on the effectiveness of lexicographic resources.
   LEXI. 9 (2), 139-174, 2022. ISSN: 2197-4292.
   DOI: http://dx.doi.org/10.1558/lexi.22164

4. **Abdelzaher, E.**: An Investigation of Corpus Contributions to Lexicographic Challenges over the Past Ten Years.
   Lexikos. 32, 162-179, 2022. ISSN: 1684-4904.
   DOI: http://dx.doi.org/10.5788/32-1-1714
   IF: 0.375 (2021)

5. **Abdelzaher, E.**: Comparing Frame Membership, WordNet-Based Similarity and Distributional Similarity.
   Computational Linguistics in the Netherlands. 10, 5-17, 2020. ISSN: 2211-4009.

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

List of other publications

Foreign language scientific articles in international journals (8)

6. **Abdelzaher, E.**: Compiling a cognition-based thematic monolingual lexicon.
   CogLS. 4 (2), 313-329, 2021. ISSN: 2213-8722.
   DOI: http://dx.doi.org/10.1075/cogls.00007.abd

7. **Abdelzaher, E.**: Lexicon-based Detection of Violence on Social Media.
   Cogn. Semant. 5 (1), 32-69, 2019. ISSN: 2352-6408.
   DOI: http://dx.doi.org/10.1163/23526416-00501002

8. **Abdelzaher, E.**: The Systematic Adaptation of Violence Contexts in the ISIS Discourse: A Contrastive Corpus-Based Study.
   Corpus Pragmatics. 3 (2), 173-203, 2019. ISSN: 2509-9507.
   DOI: http://dx.doi.org/10.1007/s41701-019-00055-y

9. **Abdelzaher, E.**, Essam, B. A.: Weaponizing words.
   JLP. 18 (6), 893-914, 2019. ISSN: 1569-2159.
   DOI: http://dx.doi.org/10.1075/jlp.18048.abd
   IF: 1.628

10. **Abdelzaher, E.**, Elghamry, K., El-Attar, A. A.: A Corpus-based Arabic Valency Dictionary: The Case of Fighting Verbs.
    Egyptian Journal of Language Engineering. 4 (1), 20-28, 2017. ISSN: 2356-8216.

11. Essam, B. A., **Abdelzaher, E.**: Approaching "lost love" theme in two culturally different poems: A cognitive linguistic analysis.
    J. Comp. Lit. Aesthet. 38 (1-2), 33-48, 2015. ISSN: 0252-8169.

12. **Abdelzaher, E.**: Semantic Framing of NATIONALISM in the National Anthems of Egypt and England.
    Int. J. Appl. Ling. Eng. Lit. 4 (4), 62-76, 2015. ISSN: 2200-3592.
    DOI: http://dx.doi.org/10.7575/aiac.ijalel.v.4n.4p.62

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

13. Essam, B. A., **Abdelzaher, E.**: Challenges in Translating Colloquial Egyptian Arabic Poetry into English: The Case of Register and Metaphors: A Contrastive Study.
International Journal of English Language and Translation Studies. 2 (3), 14-22, 2014. ISSN: 2308-5460.

Total IF of journals (all publications): 2,003
Total IF of journals (publications related to the dissertation): 0,375

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

29 March, 2023

DEBRECENI EGYETEM * Egyetemi és Nemzeti Könyvtár *

DEBRECENI EGYETEM
EGYETEMI ÉS NEMZETI KÖNYVTÁR
H-4002 Debrecen, Egyetem tér 1, Pf.: 400
Tel.: 52/410-443, e-mail: publikaciok@lib.unideb.hu

Nyilvántartási szám: DEENK/88/2023.PL
Tárgy: PhD Publikációs Lista

Jelölt: Abdelzaher, Esra
Doktori Iskola: Nyelvtudományok Doktori Iskola
MTMT azonosító: 10065286

A PhD értekezés alapjául szolgáló közlemények

Idegen nyelvű, külföldi könyvrészletek (1)

1. **Abdelzaher, E.**: Cognitive Linguistics and Digital Lexicography.
   In: The Routledge Handbook of Cognitive Linguistics / edited by Xu Wen and John R. Taylor, Routledge, New York, 568-584, 2021, (Routledge handbooks in linguistics) ISBN: 9781138490710

Idegen nyelvű tudományos közlemények hazai folyóiratban (1)

2. **Abdelzaher, E.**, Tóth, Á.: Defining Crime: A multifaceted approach based on Lexicographic Relevance and Distributional Semantics.
   Argumentum (Debr.). 16, 44-63, 2020. EISSN: 1787-3606.
   DOI: http://dx.doi.org/10.34103/ARGUMENTUM/2020/4

Idegen nyelvű tudományos közlemények külföldi folyóiratban (3)

3. **Abdelzaher, E.**: A classroom-based study on the effectiveness of lexicographic resources.
   LEXI. 9 (2), 139-174, 2022. ISSN: 2197-4292.
   DOI: http://dx.doi.org/10.1558/lexi.22164

4. **Abdelzaher, E.**: An Investigation of Corpus Contributions to Lexicographic Challenges over the Past Ten Years.
   Lexikos. 32, 162-179, 2022. ISSN: 1684-4904.
   DOI: http://dx.doi.org/10.5788/32-1-1714
   IF: 0.375 (2021)

5. **Abdelzaher, E.**: Comparing Frame Membership, WordNet-Based Similarity and Distributional Similarity.
   Computational Linguistics in the Netherlands. 10, 5-17, 2020. ISSN: 2211-4009.

DEBRECENI EGYETEM

DEBRECENI EGYETEM
EGYETEMI ÉS NEMZETI KÖNYVTÁR
H-4002 Debrecen, Egyetem tér 1, Pf.: 400
Tel.: 52/410-443, e-mail: publikaciok@lib.unideb.hu

További közlemények

Idegen nyelvű tudományos közlemények külföldi folyóiratban (8)

6. **Abdelzaher, E.**: Compiling a cognition-based thematic monolingual lexicon.
   CogLS. 4 (2), 313-329, 2021. ISSN: 2213-8722.
   DOI: http://dx.doi.org/10.1075/cogls.00007.abd

7. **Abdelzaher, E.**: Lexicon-based Detection of Violence on Social Media.
   Cogn. Semant. 5 (1), 32-69, 2019. ISSN: 2352-6408.
   DOI: http://dx.doi.org/10.1163/23526416-00501002

8. **Abdelzaher, E.**: The Systematic Adaptation of Violence Contexts in the ISIS Discourse: A Contrastive Corpus-Based Study.
   Corpus Pragmatics. 3 (2), 173-203, 2019. ISSN: 2509-9507.
   DOI: http://dx.doi.org/10.1007/s41701-019-00055-y

9. **Abdelzaher, E.**, Essam, B. A.: Weaponizing words.
   JLP. 18 (6), 893-914, 2019. ISSN: 1569-2159.
   DOI: http://dx.doi.org/10.1075/jlp.18048.abd
   IF: 1.628

10. **Abdelzaher, E.**, Elghamry, K., El-Attar, A. A.: A Corpus-based Arabic Valency Dictionary: The Case of Fighting Verbs.
    Egyptian Journal of Language Engineering. 4 (1), 20-28, 2017. ISSN: 2356-8216.

11. Essam, B. A., **Abdelzaher, E.**: Approaching "lost love" theme in two culturally different poems: A cognitive linguistic analysis.
    J. Comp. Lit. Aesthet. 38 (1-2), 33-48, 2015. ISSN: 0252-8169.

12. **Abdelzaher, E.**: Semantic Framing of NATIONALISM in the National Anthems of Egypt and England.
    Int. J. Appl. Ling. Eng. Lit. 4 (4), 62-76, 2015. ISSN: 2200-3592.
    DOI: http://dx.doi.org/10.7575/aiac.ijalel.v.4n.4p.62

DEBRECENI EGYETEM
EGYETEMI ÉS NEMZETI KÖNYVTÁR
H-4002 Debrecen, Egyetem tér 1, Pf.: 400
Tel.: 52/410-443, e-mail: publikaciok@lib.unideb.hu

13. Essam, B. A., **Abdelzaher, E.**: Challenges in Translating Colloquial Egyptian Arabic Poetry into English: The Case of Register and Metaphors: A Contrastive Study. *International Journal of English Language and Translation Studies. 2* (3), 14-22, 2014. ISSN: 2308-5460.

A közlő folyóiratok összesített impakt faktora: 2,003
A közlő folyóiratok összesített impakt faktora (az értekezés alapjául szolgáló közleményekre): 0,375

A DEENK a Jelölt által az iDEa Tudóstérbe feltöltött adatok bibliográfiai és tudománymetriai ellenőrzését a tudományos adatbázisok és a Journal Citation Reports Impact Factor lista alapján elvégezte.

Debrecen, 2023.03.29.

DEBRECENI EGYETEM * Egyetemi és Nemzeti Könyvtár *