A. S. Hornby Educational Trust

REPORT ON A.S. HORNBY DICTIONARY RESEARCH AWARD PROJECT

Title: Piloting SemiMed – a mini semantic visualization dictionary of semi-technical medical vocabulary: a response to semantic deficiencies in a medicine-related wordlist

Country: Australia/Vietnam

Dates: July 2021 – October 2022

Lead researcher: Chinh Ngan Nguyen Le

1 PROJECT SUMMARY

The project aimed to pilot SemiMed – the final product of a larger-scale project developing a mini dictionary where meanings of semi-technical medical vocabulary are visualized in semantic networks. The compilation of SemiMed stems from a demand for a reference source mainly designed for teaching/learning semi-technical medical vocabulary, because this type of vocabulary usually brings about pedagogical challenges. The starting point was Hsu's (2013) list of semi-technical medical words, whose creation and presentation incur semantic deficiencies (Le & Miller, forthcoming). Multi-meaning words in Hsu's list, which are anticipated to cause difficulties in learning and teaching, were semantically analysed with reference to theories in lexical semantics. Cantos and Sanchez's (2001) model of Lexical Constellations (LCs) was adopted as a means of showcasing intricate interrelations between general and specialized meanings of semi-technical medical words. A corpus-based analysis followed to quantify the word meaning frequency. To examine the practicality of SemiMed, a pilot study was conducted in which 18 EFL medical students were provided with lexicographical resources, including a sample of SemiMed as well as conventional dictionaries, to help them

use appropriate vocabulary while role-playing targeted medical scenarios. Focus groups were conducted to gain their feedback on the usefulness of the materials, informing improvements to SemiMed's design to better meet user needs.

2 BACKGROUND AND OBJECTIVES

2.1 Statement of Research Problem

In teaching and learning English for specific purposes (ESP), semi-technical vocabulary has long been downplayed, as greater attention has been paid to technical vocabulary. The acquisition of only technical words, however, is inadequate for a full understanding of specialized readings (Cohen et al., 1988), and the body of literature contains several studies which underscore the importance of semi-technical words and their complicated nature (Baker, 1988; Farrell, 1990; Fraser, 2009, 2012; Higgins, 1966; Le & Miller, forthcoming; Li & Pemberton, 1994). As its name suggests, semi-technical medical vocabulary is hybrid in nature, i.e., conveying general and medical meanings, and sometimes activating additional meanings in a specialized context that differ from those in the general context. By analysing 302 semi-technical medical words, Le and Miller (forthcoming) elucidate that a root cause of learning and teaching difficulties lies in polysemy and homography. Semi-technical medical words are subject to meaning variation. This type of vocabulary has multiple related (polysemic) and unrelated (homographic) meanings across different contexts and this, according to Fraser (2012), “provide[s] learners with the greatest difficulty” (p. 135)

> Semi-technical medical vocabulary is hybrid in nature, conveying general and medical meanings, and sometimes activating additional meanings in a specialized context that differ from those in the general context.

Given that semi-technical medical vocabulary has a hybrid nature, that is to say, it is found in both general and medical contexts, learners of English for medical purposes (EMP) may need to consult both general and specialized dictionaries to gain an adequate interpretation of semi-technical medical words. Moreover, most dictionaries are structured in a unidimensional format, with senses vertically listed under a dictionary word entry. This makes it harder to retrieve polysemous words, which are multidimensional in structure (Geeraerts, 2006). Therefore, this study was conducted to examine the feasibility of an alternative lexicographical

resource developed to deal exclusively with semi-technical medical vocabulary and address the semantic intricacies of polysemy and homography.

2.2 Literature Review

2.2.1 Polysemy and homography in lexical semantics

Polysemy refers to a word having multiple related meanings. Homography is a reverse phenomenon where two words with different meanings share the same written form. In lexical semantics, attention has been paid to the distinction between polysemy and homography, and the mental representation of polysemy.

There are two approaches to distinguishing polysemy from homography: etymology-based and intuitive judgment (Lyons, 1977, as cited in Klepousniotou, 2002). The former approach traces the word origin to distinguish polysemy from homography – homographs are derived from distinct roots, while polysemous words are not (Croft & Cruse, 2004; Klepousniotou, 2002). The latter approach rests on the native speaker's intuition to judge the relatedness of meaning and then determine whether meanings are closely related enough to be polysemous. Each approach has its own shortcomings, including uncertainty about the historical derivation of words (Klepousniotou, 2002) and the undesirably high level of subjectivity resulting from the existence of arbitrariness (Lyons, 1969, as cited in Atkins, 1991). Combining the two approaches may remedy the shortcomings. For example, intuitive judgment can be informed by etymology-based evidence to minimize subjectivity.

Regarding the mental representation of polysemy, it is worth mentioning that only the structural nature of polysemy is discussed within the scope of this section because polysemous meanings intertwine in a more complicated manner than homographs and thus need elaboration. Two standpoints that merit discussion in the study context are Ruhl's (1989) *monosemy* and Lakoff's (1987) *radial category*. Ruhl (1989, 2002) argues that despite having many meanings, only one abstract meaning is stored in the brain; other meanings of a polysemous word are constructed via semantic and pragmatic context clues. By contrast, Lakoff (1987) maintains that a polysemous word is a conceptual category and we store "a category of distinct polysemous senses rather than a single abstract monosemous sense" (as cited in Evans & Green, 2006, p. 330).

The development of Lakoff's proposed radial categories was later parameterized by Tyler and Evans's (2003) principles (also known as the *principled polysemy* approach). However, in

essence, a radial category does not change its nature. It remains “a conceptual category in which the range of concepts are organised relative to a central or prototypical concept” (Evans & Green, 2006, p. 331). In other words, the radial category visualizes how different meanings of a word interact vis-à-vis a central meaning, the one that typically presents mutual semantic properties of other meanings. Polysemy under this perspective is structured in a “highly complex” way (Evans & Green, 2006, p. 328) and has “multidimensional structural relations” (Geeraerts, 2006, p. 351).

2.2.2 Polysemy and homography in lexicography

Turning now to polysemy and homography from the perspective of lexicography, ways in which these two phenomena are handled in two lexicographical resources (wordlists and dictionaries) will be discussed.

Wordlists

A wordlist is a list that indicates a finite number of words learners need to master for their particular learning purposes. For example, Hsu’s (2013) Medical Word List (MWL) contains 595 semi-technical words that appear so frequently that learners of EMP are advised to spend their time learning the listed words to gain adequate comprehension of what they hear or read. Having the frequency of word forms as an underlying basis for the selection of candidate words, frequency wordlists come at a price, that is, the wordlist creation and presentation do not pay due attention to semantic relations. The wordlist creation rests upon the automatic corpus-based distinction of word forms rather than word meanings, thereby disregarding the phenomena of polysemy and homography (Watson-Todd, 2017), and consequently failing to include them in the presentation. Although wordlists play a significant role in delimiting vocabulary size and thus letting learners know which words they should focus on, Le and Miller (forthcoming) express a growing concern over the absence of semantic explanation in wordlists, especially wordlists of semi-technical words like the MWL, in which 51% of words are polysemes or homographs or both.

Dictionaries

Compared with wordlists, dictionaries have more sufficient and elaborated presentation of polysemy and homography. There are several ways to order related and unrelated meanings within a dictionary entry. Still, for reasons of space, only two internal structures are discussed in this report because they are the ones most commonly used in conventional dictionaries. These are *linearization* and *hierarchy*.

In a linear structure, “all [meanings] have equal status … [and] are presented on one level” (Moerdijk, 2003, p. 285). A hierarchical structure, on the other hand, has “two or more levels on which related [meanings] are grouped” (p. 286). It has been argued, however, that these internal structures do not fully capture the semantic intricacies of polysemy and homography. Given that meanings are all listed on the same level, linearization may not imply semantic inter-relatedness and thus, dictionary users, when they look at linearly organized meanings, may tend to treat each meaning as a discrete element that has no relation to remaining meanings. For example, Cambridge Dictionary (https://dictionary.cambridge.org/) does not flag (a) the distinction between homography and polysemy, and (b) the relation among polysemous meanings due to its linear structure of presenting homographs (e.g., *colon*) and polysemous meanings of words (e.g., *benign*) at the same level (Figure 1).

colon	**benign**
n. (body part) the lower and bigger half of the bowels in which water is removed from solid waste n. (sign) the symbol: used in writing, especially to introduce a list of things or a sentence or phrase taken from somewhere else	adj. (person) pleasant and kind adj. (disease) a benign growth is not cancer and is not likely to be harmful

Figure 1: The linear structure of *colon* and *benign*. Definitions from Cambridge Dictionary online (https://dictionary.cambridge.org/) in the order in which they appear.

In a hierarchical structure, by contrast, meaning groupings determined from their relatedness seem to provide more straightforward indications than the linear structure does. The distinction between homography and polysemy is drawn because homographs and polysemous words are grouped in separate entries. Polysemous meanings are grouped within an entry in a hierarchical order (Figure 2). Nevertheless, although the hierarchy of polysemous meanings establishes their relation, how the different meanings relate to each other is not explicitly showcased. In other words, from the standpoint of lexical semantics, particularly the multidimensional structural relations of polysemy (Geeraerts, 2006), a hierarchical structure still has a minimal capacity for showcasing polysemous relations.

colon (n)	**benign** (adj)
Entry 1: the part of the large intestine that extends from the cecum to the rectum	1a: of a mild type or character that does not threaten health or life
Entry 2	b: having no significant effect
1 plural cola: a rhythmical unit of an utterance	2: of a gentle disposition
2 plural colons:	3a: showing kindness and gentleness
a: a punctuation mark	b: favourable, wholesome
b: the sign	
Entry 3: a colonial farmer or plantation owner	

Figure 2: The hierarchical structure of *colon* and *benign*. Definitions from Merriam-Webster Dictionary online (https://www.merriam-webster.com/) in the order in which they appear.

> This raises questions as to whether the conventional format of lexical resources does full justice to the intractable nature of linguistic phenomena.

3 DESCRIPTION OF RESEARCH

A review of the literature, then, indicates that the semantic structures in lexical semantics are not fully observable in lexicographical practice, raising questions as to whether the conventional format of lexical resources does full justice to the intractable nature of linguistic phenomena. To begin to address this issue, this study aimed to develop a non-conventional lexical resource of semi-technical medical vocabulary that takes into account theories of polysemy and homography in lexical semantics. The study had two phases:

- Developing a pilot version of SemiMed, an exclusive resource of semi-technical medical vocabulary that considers polysemy and homography from the perspective of lexical semantics
- Piloting SemiMed to test its usefulness in comparison with current conventional resources

3.1 Developing SemiMed

The MWL was a starting point for the development of SemiMed. The study took advantage of Le and Miller's (forthcoming) findings and conducted a semantic analysis of 302 multi-meaning semi-technical medical words in the MWL. The MWL was chosen for two reasons:

- A wordlist, unlike a dictionary, usually has a finite number of words. This would ensure the feasibility of the study. More importantly, although the MWL has semantic issues due to its reliance on word form frequency, it still informs us of semi-technical words that frequently occur in medical contexts.
- A semantic analysis of words in a wordlist is more pedagogically significant than analysis of words in a dictionary. Words in the MWL are chosen selectively on the basis of frequency, which means they occur so frequently in medicine that EMP learners should devote time and effort to learning them. In comparison, not every word in a general/medical dictionary is worth learning. Additionally, a wordlist has minimum semantic features, so the semantic improvement of the MWL may be expected to compensate for the shortcomings of word form-based wordlists and so pave the way for the development of a resource containing frequently occurring semi-technical medical words with sufficient semantic explanation.

3.1.1 Qualitative analysis

The qualitative analysis of 302 MWL words was rooted in the theories of lexical semantics reviewed above. First, the analysis used a combined approach that considered both etymology and speaker intuition to distinguish polysemy from homography. Second, although Ruhl (1989) and Lakoff (1987) hold contradictory views on how a polysemous word is mentally stored, at the heart of monosemy and radial category, a shared concept can be observed of a core meaning (variously named an abstract, central or prototypical meaning) – the one from which polysemous meanings are derived. Following this observation, a visualization of polysemous relations was proposed in response to the hierarchical structure's minimal capacity to showcase how each polysemous meaning interrelates with others. This allows a higher level of hierarchical structure, where polysemous relations are not implicit or implied but explicitly visualized. Rather than vertically listing polysemous meanings under a word entry such as *benign* in Figure 2, the qualitative analysis further visualizes how polysemous meanings interact vis-à-vis a core meaning. The highly complex structure of polysemy in Lakoff's radial category is acknowledged, and his idea that "the range of concepts are organized relative to a central or prototypical concept" (Evans & Green, 2006, p. 331) helps to explain the semantic visualization.

3.1.2 Quantitative analysis

To address the word form frequency-related issues, a corpus-based analysis was carried out to examine how frequently word meanings presented in our semantic visualization appear in a general and a medical corpus. Two corpora were selected – English Web 2020 (36 billion words) and the Medical Web Corpus (34 million words) (Table 1). The unit of analysis was word meaning frequency and the unit of counting was collocate frequency. The analytical method was based on an approach to determining meaning by collocation (Cantos, Sanchez, & Almela, 2009; Hoey, 2012; Perez, 2013). Simply put, the meaning interpretation of a word in a corpus is retrieved from an extensive investigation into its collocations. The collocational data were computed using the online corpus analysis tool Sketch Engine.

Table 1: Details of English Web 2020 and Medical Web Corpus

	English Web 2020 (enTenTen20)	**Medical Web Corpus**
Tokens	43,125,207,462	42,054,011
Words	36,561,273,153	33,961,786
Sentences	2,008,143,278	1,545,862
Documents	78,373,887	526

3.1.3 Procedural demonstration of *diffuse*

Step 1: Oxford English Dictionary (OED) definition adaptation and simplification

An MWL headword (e.g., *diffuse*, see Figure 4) was prepared by adapting the procedure for looking up MWL headwords in the OED used by Le and Miller (forthcoming). OED definitions were then simplified to:

- Make OED definitions shorter and easier to understand for learners at a minimum upper-intermediate level of English proficiency
- Ensure the use of simply reworded OED definitions in the semantic visualization does not infringe copyright

Step 2:

Identification of core and other related meanings

The principled polysemy approach informed the identification of core and other related meanings. Criteria to determine a core meaning in this study were derived from Evans (2005). A core meaning needs to fulfil at least one, and preferably more than one, of three criteria: "(1) [closely relates to the] historically earliest attested meaning, (2) predominance in the semantic

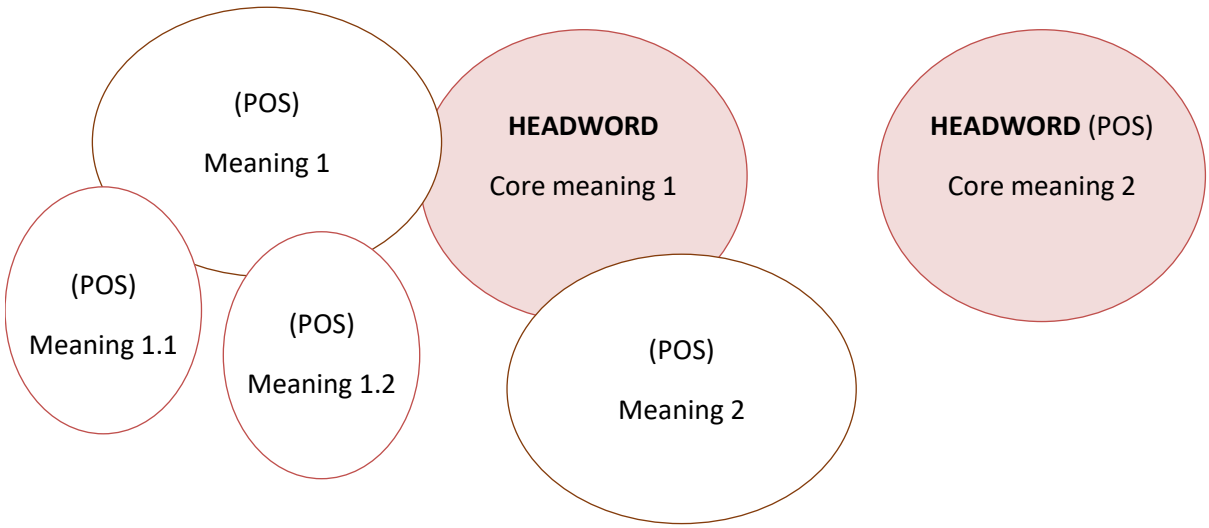
network, […] (3) predictability regarding other senses" (Evans, 2005, p. 44). The study adopted Le and Miller's (forthcoming) identified core meanings of 302 troublesome semi-technical medical words from the MWL. Briefly, Le and Miller evaluated the first criterion by looking at the etymological reference in the OED. The remaining criteria were based on the intuitive judgment of three evaluators. From the core meaning identification using this combined approach, Le and Miller reasoned that they would classify a word as polysemous if all its senses shared a core meaning; new senses creating new core meanings would be classified as homographs. This reasoning was used to distinguish polysemy from homography.

Hierarchy of other related meanings

Non-core meanings were further analysed by putting closely related meanings into a cluster and establishing a meaning hierarchy within a cluster.

Step 3: Visualization of semantic relations in Lexical Constellations

The study adapted Cantos and Sanchez's (2001) Lexical Constellation (LC) model to visualize how related meanings interact vis-à-vis a core meaning. The generic pattern of an LC has a core meaning placed at the centre and surrounded by multiple, related meanings located in outer layers, which showcase the degree of interconnectivity (Figure 3). Each LC represents a (polysemous) word and if two words are homographs, they have two separate LCs.



POS: Part of Speech
Level 1: Meanings 1 and 2
Level 2: Meanings 1.1 and 1.2

Figure 3: Generic pattern of LCs (of a polysemous word and a homograph).
Adapted from Rizzo & Sanchez (2010).

Step 4: Quantification of the meaning frequency of occurrence

Sketch Engine (Word Sketch) was used to export collocates of a searched headword and select the top 15 most frequent collocates in two corpora (English Web 2020 and the Medical Web Corpus). Meanings were assigned to collocates and then divided into four levels of technicality (Table 2). Level 0 indicated that no technicality information was shown for a meaning, the meaning is not found in the Medical Web Corpus, and it is considered a purely general meaning. Technicality levels 1 – 3 were embedded in LCs (Step 4 in Figure 4).

> Sketch Engine was used to select the top 15 most frequent collocates in two corpora; one general and one medical corpus. Meanings were assigned to collocates, and then divided into four levels of technicality.

Table 2: Technicality level description

Level 0 Not indicated in LCs	**This meaning is solely used in general contexts** Found in the top 15 meanings in English Web 2020, but not found in the Medical Web Corpus
Level 1	**This is a generally used meaning** Found in the top 15 meanings in English Web 2020, and outside of the top 15 meanings for the Medical Web Corpus
Level 2	**This meaning is used in both general and medical contexts** Found in the top 15 meanings in both English Web 2020 and the Medical Web Corpus
Level 3	**This meaning is used only in medical contexts** Found in the top 15 meanings in the Medical Web Corpus, and outside of the top 15 meanings for English Web 2020; or Found in the top 15 meanings in the Medical Web Corpus, but not found in English Web 2020

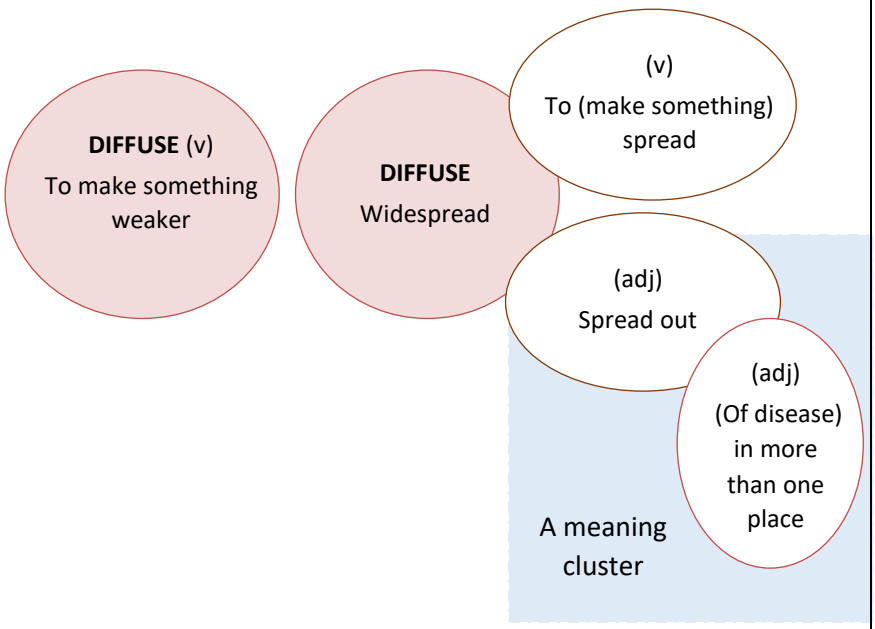
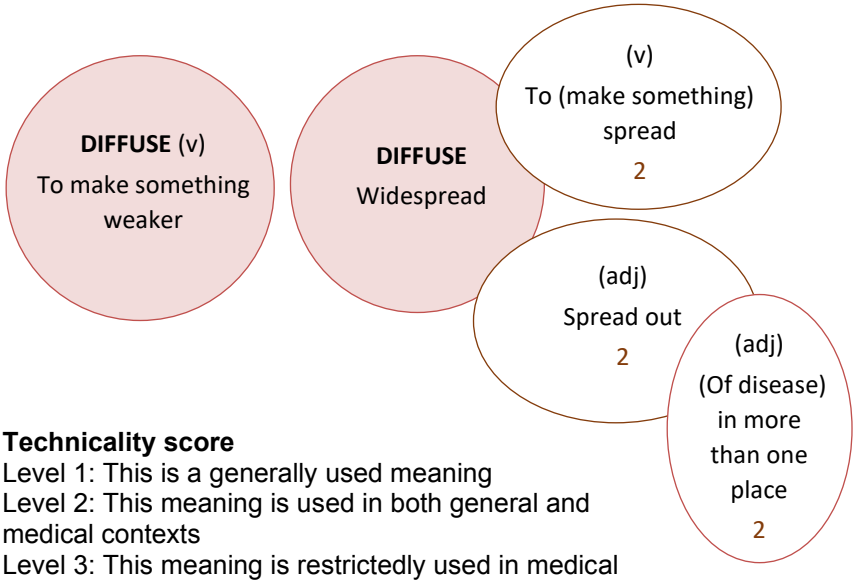
Step 1: Simplify OED definitions of *diffuse*	**DIFFUSE** (adj) Spread out (adj) (Of disease) in more than one place (v) To (make something) spread (v) To make something weaker
Step 2.1: Identify 2 core meanings and 3 other meanings relating to core meaning 1 **Step 2.2:** Put the 3 other meanings in clusters and indicate the hierarchy	**DIFFUSE** ***Core meaning 1: Widespread*** Meaning 1: (adj) Spread out Meaning 1.1: (adj) (Of disease) in more than one place Meaning 2: (v) To (make something) spread ***Core meaning 2: To make something weaker***
Step 3: Develop 2 LCs of *diffuse* (with 3 polysemous meanings) and its homograph	
Step 4: Examine the meaning frequency of the polysemous word and homograph	 **Technicality score** Level 1: This is a generally used meaning Level 2: This meaning is used in both general and medical contexts Level 3: This meaning is restrictedly used in medical contexts

Figure 4: Procedural demonstration of *diffuse*

3.2 Piloting SemiMed

3.2.1 Sampling

Forty LCs were selected for the pilot study. A wide range of LC constructs were taken into consideration during the sampling process to ensure pilot words closely reflected the characteristics of SemiMed LCs. The sample included LCs of (a) single-meaning words (e.g., *colon*, Figure 5), (b) multi-meaning words with a single core meaning (e.g., *benign*, Figure 6) and (c) multi-meaning words with more than one core meaning (e.g., *diffuse*, Figure 7).

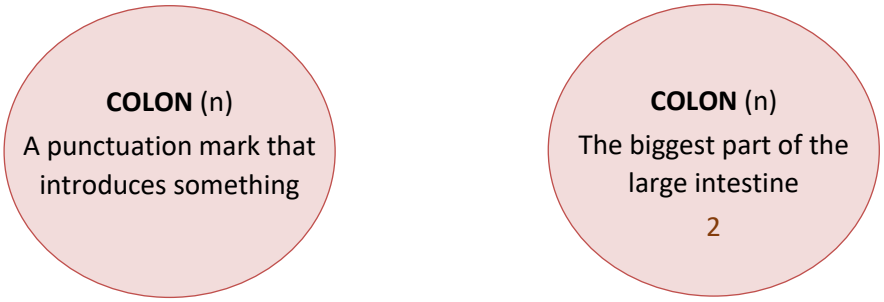


Figure 5: Two homographs

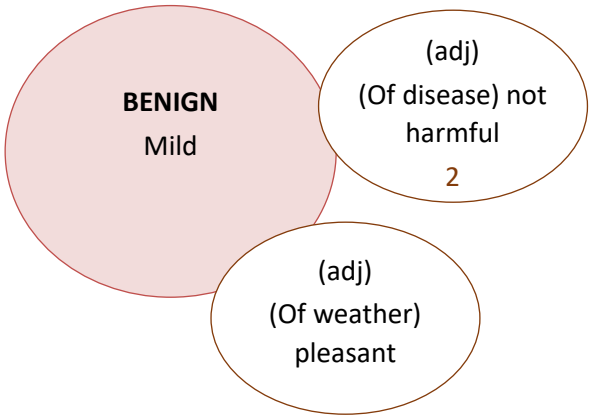


Figure 6: A polysemous word

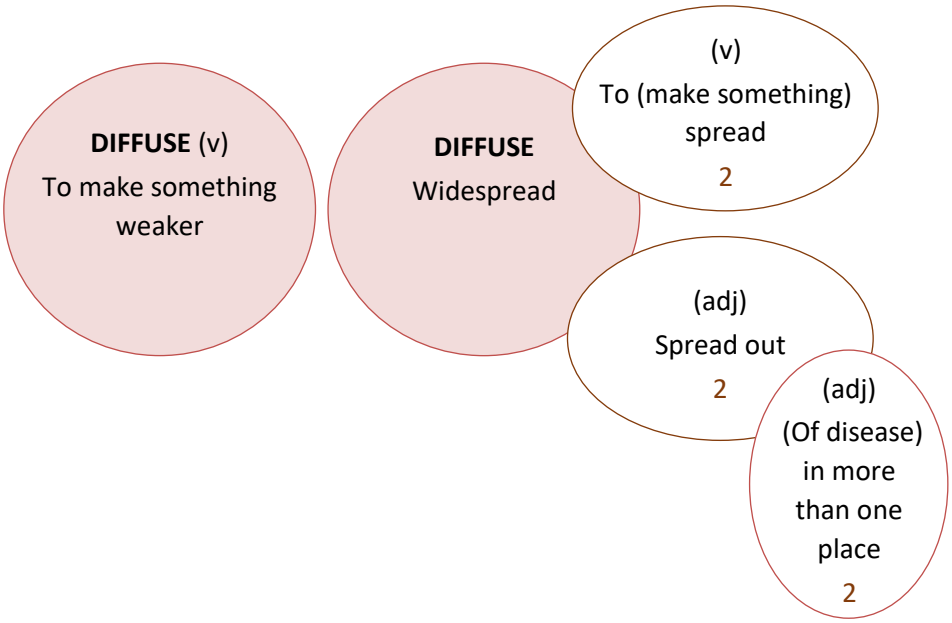


Figure 7: A polysemous word and a homograph

3.2.2 Participants

Eighteen EFL medical students from a University of Medicine and Pharmacy (UMP) in Vietnam were participants in the pilot study. They were recruited based on their English proficiency. Eligible participants were students who majored in medical fields and possessed an upper-intermediate or higher level of English.

3.2.3 Lexicographical resources

Participants were allowed to use three resources:

SemiMed which presents word meanings in the format of an LC.

Two designated dictionaries

- A general dictionary: Cambridge Dictionary (https://dictionary.cambridge.org/), in which definitions are presented in the linear format.
- A specialized dictionary: Merriam-Webster Medical Dictionary (https://www.merriam-webster.com/medical), in which definitions are presented in the hierarchical format.

3.2.4 Online platform

All forty pilot LCs were drawn using Inkscape software, then uploaded onto H5P (https://h5p.org/) and finally embedded in the UMP's Moodle system for participants to access. The LCs were alphabetically ordered and presented in four 'books' for ease of access

(SemiMed A-C, SemiMed D-I, SemiMed L-P, and SemiMed R-T) (Figure 8). A pop-up box was designed to show detailed information of the technicality level (Figure 9).

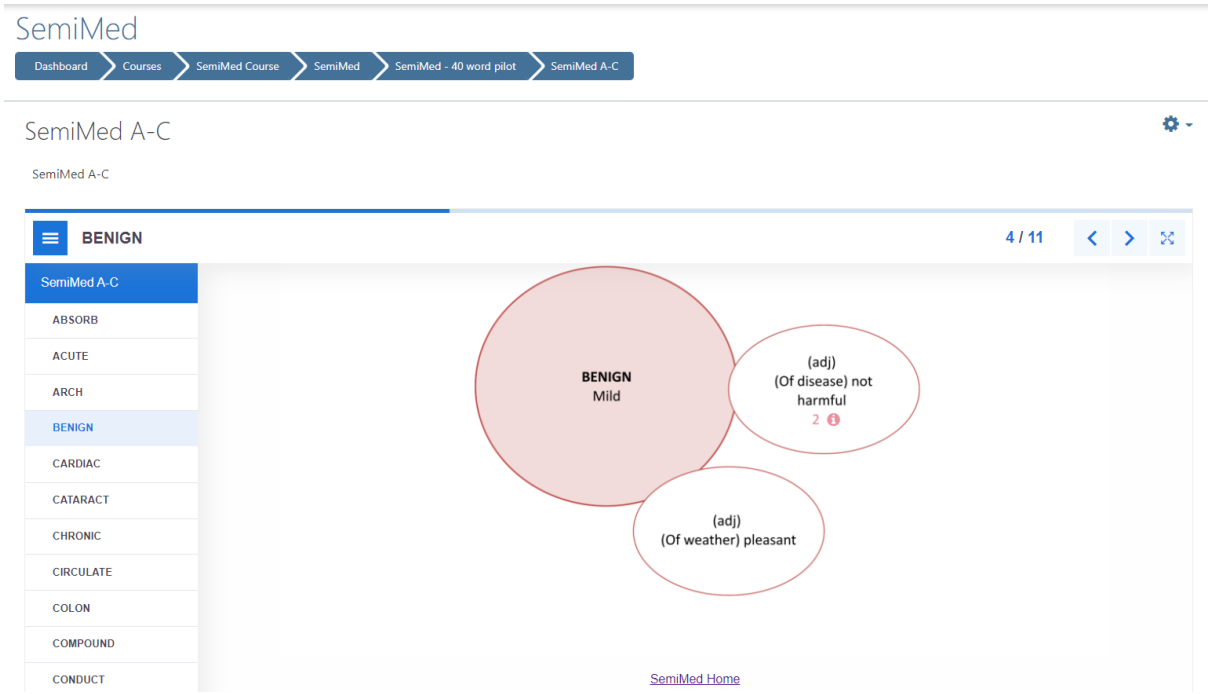


Figure 8: A Moodle interface of the LC of *Benign* (in Book 1: SemiMed A-C)

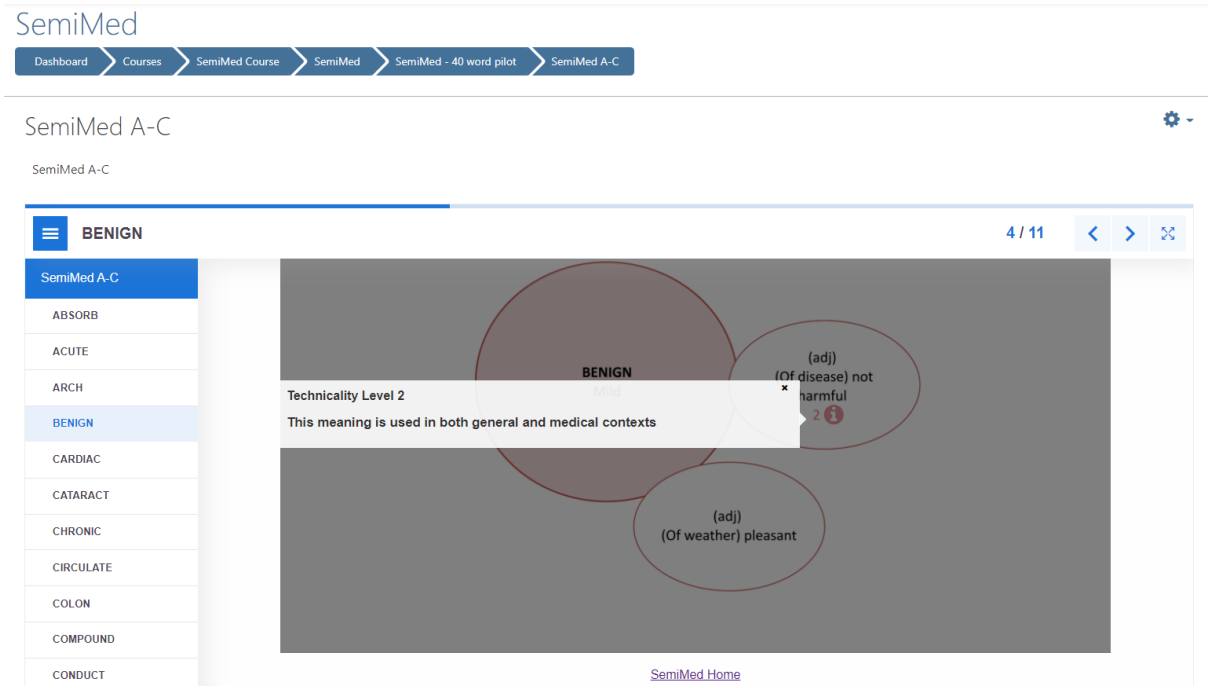


Figure 9: A pop-up box indicating the detailed technicality level

3.2.5 Medical scenarios

Five scenarios were written around topics that closely mimicked real-life situations that participants might experience. Each scenario targeted six pilot words, making a total of 30. The remaining ten words were example words shown to participants in the Induction phase of the study (see the Meeting structure). In essence, scenarios set the scene to stimulate participants to use the pilot words in meaningful and relevant contexts. Gaps, indicated by ellipses, were left in the scripts to prompt participants to explain the target words to the 'patient' in their role play.

3.2.6 Grouping

Eighteen participants were randomly divided into six groups (three people per group). The researchers scheduled a separate online Zoom meeting with each group.

3.2.7 Meeting structure

Induction: Participants were introduced to SemiMed and instructed to use this new resource, especially to interpret information presented in LCs. Participants were also informed of the dictionaries they were requested to use.

Activities: Participants chose their roles in scenarios and acted out the scenarios. They were encouraged to consult the lexicographical resources provided to use the pilot words as appropriately as possible. Specifically, they were requested to look up the first pilot word in the Cambridge Dictionary online, the second one in Merriam-Webster Dictionary online, and the third one in SemiMed. The rest of the pilot words could be looked up in any dictionary, allowing participants to choose which format they preferred. The researchers observed and facilitated as needed.

Focus group: Participants then engaged in a follow-up focus group where they shared their experiences of using SemiMed and the conventional dictionaries.

4 RESULTS AND EVALUATION

Thematic analysis was used to analyse the focus group data. The themes which emerged centred around participants' experiences of using SemiMed in the pilot study and also extended to their experiences of using other conventional resources prior to the pilot study.

4.1 Conventional resources

The majority of participants reported that, before the study, their two most frequently used general dictionaries were those published by Cambridge and Oxford (titles and editions were not given). In addition to monolingual dictionaries, they sometimes referred to bilingual dictionaries (SOHA[1] and TFLAT[2]) to search for Vietnamese meanings. Surprisingly, they seldom used medical dictionaries. Some mentioned *The language of medicine* (Chabner, 2020) as the only resource formally introduced in classrooms that provided them with topic-based medical terminology and learning strategies (e.g., morphemic analysis). Many participants used Google Search and Google Translate, which according to Participant O was a strategy passed down from senior to freshman students. Participants exploited these two functionalities of Google in various ways, ranging from looking up words to checking meanings of a known word.

Participant O shared that she usually put *what does word X mean?* in the search box and emphasized that "a strength of Google [Search] is that it provides you with images and some kinds of videos so that it helps you understand the word more clearly". Several participants also considered Google Search engine as a medium for seeking related visual aids to assist them in understanding and learning a word. Participant B, for example, stated:

> *I think the most problem I get when I try to find meanings of the English medical terms is that there are some rare medical words I don't find on the Internet so I have to look up [a word] on the Google Images and I see the picture of it and I will have to try to guess [its] meaning.*

Another student went straight to video searching:

> *I prefer Youtube [videos] so I can learn more about [a] medical word.* (Participant D)

Another common strategy shared among Participants B, C and M was doing Google searches for articles containing a specific word. They revealed that the retrieved articles offered contextual clues by which they could guess a meaning of the word. Google Translate was also used to get an instant Vietnamese translation of an English article (Participant C) or the Vietnamese equivalent of an unknown word (Participants A, N and R). Participant F used Google Translate for "fast-checking" whether she had correctly understood a word definition in the Oxford dictionary.

---

[1] An e-dictionary available at http://tratu.soha.vn/dict/en_vn/Dictionary

[2] An English Dictionary App developed by TFLAT, a mobile application development team based in Vietnam

Besides looking for and checking meanings of a word, Participant H added that she searched for the etymology of a word via Google. She also took advantage of Google to further learn about roots, prefixes, or suffixes from which a searched word is built. For Participant H, knowing the constituent parts of a searched word somehow made possible the guessing of the meaning of a new word made up of the same parts.

As previously stated, most participants reported little experience of using medical dictionaries. There are many possible reasons for this. Participant H reasoned that although she had been informed about medical dictionaries, she had never used one, as she could not afford the subscription fee. Participant K admitted that "I am afraid of being not fully understand [sic] words [in medical dictionaries]". Participant R had used medical dictionaries but thought that the definitions of a word were sometimes more complex than the word itself. Participant M asserted that he had no intention of finding a medical dictionary:

> *I am a visual learner so I think that for medical dictionaries just [containing] words, they are not just for me.*

Participants appeared to rely heavily on general dictionaries to look for medical meanings. However, they reported low satisfaction with the use of general dictionaries because they did not always find what they were seeking.

> *When I look up the meaning [of a word] in the [Cambridge and Oxford] dictionaries, they normally show the general use of the word and sometimes that word doesn't have the … sometimes I cannot find the technical meaning.* (Participant C)
>
> *For me, when I [try to] find some medical words in Cambridge or Oxford [dictionaries], there is no result so I have to use Google to find the meanings of the medical words that I want to figure out.* (Participant I)
>
> *When I [used] Cambridge dictionary, some of technical words didn't appear.* (Participant K)

The possible inference of this feedback is that non-specialized Cambridge and Oxford dictionaries are not ideal for searching for medical meanings. Moreover, Participant N commented that the two general dictionaries occasionally led to homographs irrelevant to the medical context and this distracted her. In the case of medical meanings found in the general dictionaries, Participant O revealed that she found definitions in the Oxford dictionary too lengthy to arrive at appropriate Vietnamese equivalents.

The challenges faced by participants while searching for medical meanings in general dictionaries may have contributed to their preference for *Google Search and Translate*.

Nevertheless, the Google tools raise some concerns. Participant F admitted that although Google Search helped her target relevant medical articles or books, it was relatively time-consuming to understand a page of the books or even a paragraph of the articles where a searched word appeared. Worse still, she sometimes failed to double-check meaning(s) as they still did not make sense to her after reading through the translated parts. Participant L recalled that he sometimes had to read up to three documents but could not work out word meanings by himself, so he eventually consulted his teachers. Participant Q was concerned that understanding a segment of articles (or books) retrieved from Google Search was exhausting because she needed to do other searches to be sure that she fully understood the entire segment containing the searched word.

The participants' experiences of conventional resources may highlight the concerning issue mentioned at the start of this report regarding the search for semi-technical vocabulary in general and specialized dictionaries. Participants seemed to get limited benefits from general (Cambridge and Oxford dictionaries) and medical (*The language of medicine*) dictionaries and this eventually drove them to use Google tools which caused them even more trouble. This finding supports the early stated need to develop a lexical resource of semi-technical vocabulary with an aim of creating a better experience in looking up this type of vocabulary.

4.2 SemiMed

SemiMed was developed to serve the practical need for a semi-technical vocabulary resource with dual foci on the logical presentation of general and medical meanings, and explicit guidance on polysemy and homography. In the pilot study, feedback from participants on the usefulness of SemiMed compared to the designated dictionaries was expected. However, the focus group data uncovered that participants also reflected on resources they used beyond the pilot study (as listed in the above section) and compared them with SemiMed. This newly emerging theme intertwined with the expected theme and provides a much better insight into the usefulness of SemiMed in comparison with current conventional resources; therefore, the two themes have been reported simultaneously rather than separately.

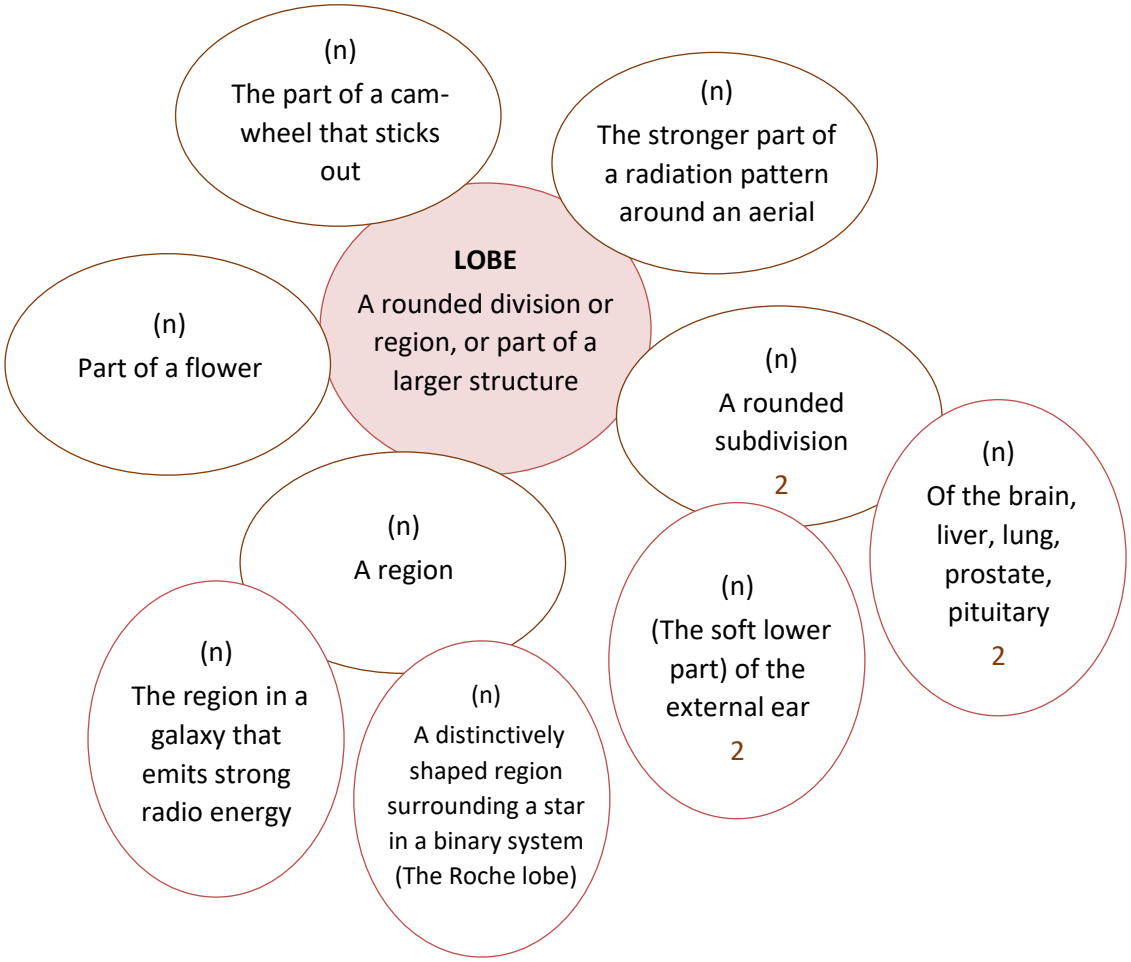
Participants identified three main advantages of SemiMed over conventional resources.

1 *Concise and simplified definitions*

Participants O and M gave feedback on the way word definitions are written in SemiMed compared to the two designated dictionaries. Participant O said when she looked up the last three pilot words in the Cambridge dictionary, she found their definitions lengthy. She thus anticipated that if she had acted out her role as a specialist and explained the pilot words using definitions from the Cambridge dictionary, the group member who played the role of a patient

might have become confused. When she looked up the same pilot words in SemiMed, she said that "the word is explained in a very short and simplified way, so I suppose that [SemiMed] will be applicable in real context when we have [a] conversation with our patients". Participant M agreed and gave the pilot word *lobe* as an example of a word being more concisely defined in SemiMed than in the Cambridge or Merriam-Webster dictionaries (see Figure 10).

SemiMed dictionary



Cambridge Dictionary

lobe

n. (anatomy) any part of an organ that seems to be separate in some way from the rest, especially one of the parts of the brain, lungs, or liver

n. (ear) an earlobe

n. (biology) a rounded or pointed part on a leaf that sticks out from the main part

Merriam-Webster Medical Dictionary

lobe (n)

A curved or rounded projection or division: as

a: a more or less rounded projection of a body organ or part

b: a division of a body organ (as the brain, lungs, or liver) marked off by a fissure on the surface

Figure 10: *Lobe* in the three dictionary formats

2 *Non-conventional format*

The participants' overview of the SemiMed format was that it is systematic, neat, and simple. Several participants (B, C, D, J and O) perceived SemiMed as a "mind map", with which medical students, according to Participant C, are familiar because they usually use mind maps for lesson revision. Participant B was impressed that SemiMed followed a mind map-like design to present word meanings and this design permitted the systematic learning of semi-technical medical words. Participant Q, agreeing with Participant B, reasoned that the inclusion of both general and medical meanings helped her to form "a general view of all meanings". This is considered a prominent advantage over conventional resources participants had used, such as Cambridge or Oxford dictionaries and Google tools, because four participants (B, C, F and J) could retrieve and understand general and medical meanings in SemiMed using only a single search.

> One participant pinpointed that SemiMed laid out meaning interrelations in a logical manner, facilitating his ability to see how general meanings interact with medical ones.

Moreover, Participant A pinpointed that SemiMed laid out meaning interrelations in a logical manner, facilitating his ability to see how general meanings interact with medical ones. Participant H said that by knowing the relationship between general and medical meanings, she might expand her knowledge about the general meanings through the learning of medical meanings and vice versa. In the role-play activity, the systematic visualization of relationships between general and medical meanings might have created a better experience compared to the two designated dictionaries; many participants reported that it was more convenient (Participant I), more helpful (Participant N), and faster (Participant Q) to search for and find appropriate medical meanings of pilot words in SemiMed than in the Cambridge and Merriam-Webster dictionaries.

Participants acknowledged the explicit distinction between polysemous words and homographs, which Participant I admitted he had not observed in the Cambridge and Oxford dictionaries. The feedback from participants revealed that the polysemy and homography presentation in SemiMed was easy to understand (Participants G, M and P) and more importantly, led them to the medical meanings of a word they were looking for, not the

irrelevant meanings of its homographs (Participants E and F). In addition to reducing the distraction from unwanted homographs, the non-linear format, particularly the radial visualization, assisted participants in learning polysemous meanings of a word. Participant D commented that it was good to know a core meaning was shared among related meanings, and reiterated that if he understood the core meanings the mind map-like structure would enable him to memorize polysemous meanings faster.

Another significant advantage of SemiMed is its clear and neat display. Participants D, J and K liked the fact that SemiMed used a one-page display view, which helped them stay focused. Participant D stated that information in the Cambridge and Merriam-Webster dictionaries was so detailed that he sometimes lost his focus. Additionally, because they show word definitions in the form of linear lists, these dictionaries require participants to scroll up and down to read through search results. Participants J and L explained that it was fairly time-consuming to scroll through the entries to find the meanings used in the scenarios. In contrast, they felt SemiMed saved considerable time as the semantic visualization of a word was designed to fit the screen. In other words, participants were likely to spend less time manipulating displayed contents and thus their focus on finding meanings was enhanced.

Although SemiMed had a non-conventional format, no participants reported challenges in familiarizing themselves with it. Rather, the findings showed consensus among participants, emphasizing that the SemiMed interface is simple and user-friendly. Participant P clarified:

> *a strength [of SemiMed] is its format … it's simple and clear … so it's kind of easy to understand … suitable for beginners and when people use it, we don't need to [have] a lot of technical and literacy skills.*

For this reason, Participant K said she could manipulate SemiMed with ease after being guided through its functions in the Induction. The easy-to-use design seems to offer participants quick access to pilot words (Participant I) and then provide scaffolding for their understanding of word meanings (Participant E).

3 *Technicality level*

The level of technicality is a feature peculiar to SemiMed which attracted positive feedback from participants. One benefit of the technicality level is that it informs users of the context in which a certain meaning is more likely to appear. As explained earlier, semi-technical medical words can be used across different contexts, so the contextual details provided for each meaning are important (Participants J and O). Since medical meanings are central to learning

semi-technical medical words, participants mainly commented on how the technicality level facilitated their search for medical meanings. When relating the difficulties in finding medical meanings using conventional general dictionaries (Cambridge and Oxford), Participant C shared that he had a more pleasant experience using SemiMed, especially its technicality function. Participants A and R were impressed by the technicality of meanings, which made the search for pilot words used in the medical scenarios much easier than searches in the designated dictionaries (Cambridge and Merriam-Webster). Furthermore, Participants C, F, and L agreed that the technicality level information significantly reduced the time allocated to searching for medical meanings. Thanks to the technicality information, Participant C was certain he spent less time finding medical meanings of pilot words, Participant F said she could know immediately which meanings fitted in medical scenarios, and Participant L stated that his focus was quickly directed to medical meanings.

> Several participants agreed that the technicality level information significantly reduced the time allocated to searching for medical meanings.

The three advantages of SemiMed are related to features absent in conventional dictionaries. These results are likely to further support the idea of developing SemiMed. Three implications were accordingly drawn from the findings.

First, participants positively reacted to the simplified definitions in SemiMed and this finding underscores the importance of well-written word definitions in medical dictionaries. Here, a “well-written” definition is understood to be one that has been constructed so that it is as easy as possible to understand by learners at all language levels. Although issues around writing a definition of a word have long been situated at the heart of the dictionary-making procedure, it is still believed that the issues deserve more attention, especially in the compilation of medical dictionaries, as this study indicated that difficulties in understanding definitions may make learners hesitant to use medical dictionaries.

Second, the radial structure seems more advantageous than a hierarchical format in terms of leveraging insight into relations (i.e., polysemy and homography) between general and medical meanings. SemiMed's non-conventional format, which adheres to theories in lexical semantics, is de facto the mental representation of polysemy and homography. That may explain why participants considered the SemiMed format beneficial in facilitating mental

processes such as understanding and memorizing general and medical meanings of semi-technical vocabulary. This implication provides some support for the consideration of lexical semantic theories in the development of lexical resources.

Third, participants' appreciation of the technicality function indicates that word meaning frequency results can be transferrable into the four technicality levels, enhancing the pedagogical usefulness of SemiMed and showing that improvement of word form frequency-based wordlists can be achieved. This has potential for resolving word form frequency-related issues in the MWL and other wordlists.

Despite advantages in word definitions, presentation format and technicality function, however, SemiMed nevertheless has some disadvantages.

Time-efficient platform manipulation and not-so-attractive interface

As explained previously, SemiMed was uploaded onto H5P and this online platform supported four 'books' of alphabetically ordered words (SemiMed A-C, D-I, L-P and R-T). Technically speaking, to look up *benign*, for example, participants had to access the first book (SemiMed A-C) and scroll down until they retrieved the word (see Figure 8). This manual method of looking up a new word, which closely mimics the traditional method used with paper-based dictionaries, created a little confusion for participants. Participant P said that even though he kept a searched word in mind, he sometimes lost his train of thought and couldn't decide which book he should select to find the word. He admitted to singing the ABC song to himself to aid his memory. Participant P added:

> *the weakness [of SemiMed] is that it has no finding tool so maybe sometimes it's very time-consuming when I have to scroll down and search for the word.*

Agreeing with Participant P, Participants I, J and L reported that this manual search of pilot words in SemiMed took more time than when using the two designated dictionaries.

Regarding SemiMed's interface, the minimal design was intentionally chosen, and this was evaluated as simple and user-friendly by many participants. However, a few participants (B, F and R) still viewed the SemiMed design as less attractive and would have liked to see additional visual features. Participants K, L and R added that the absence of illustrative pictures in SemiMed not only made the interface look monotonous but also meant that new words could not be learnt by looking at pictures.

Insufficient pronunciation and examples

Another disadvantage of SemiMed is the paucity of pronunciation guidance and of examples. Participants A, I, K and Q expressed their need to know how to pronounce a word in addition to its meanings. Participant Q reasoned that maybe because SemiMed did not show her how to pronounce a word, it was not of much benefit when speaking. In addition, Participant N said the lack of examples stopped SemiMed users from seeing a word in context. Participant E, when comparing SemiMed with the conventional dictionaries she used, stated that Cambridge and Oxford dictionaries gave her examples which enabled her to better understand what a word meant and how to use it. Participant Q added that some SemiMed word definitions were too concise to be readily understandable, so it was difficult to gain an adequate understanding by reading definitions with no examples. For instance, she could not adequately understand *conduct*, whose definition was *to do* in SemiMed, until she searched for relevant examples in the Cambridge dictionary.

4.3 Suggestions for future improvement

Suggestions were made around potential features which participants believed should be added to SemiMed to mitigate its current disadvantages. First, Participants J, L and P suggested that the online platform should be upgraded with a search bar to automate the word searching process. Rather than manually looking up a new word in the four books, typing the word in the search bar and then clicking a search button to retrieve search results seemed to be more time-efficient and thus might create a more pleasant experience for users. Second, to maximize benefits, many participants recommended the incorporation of pronunciation aspects into SemiMed so that they could both read and pronounce a word correctly. Third, they recommended that images and pictures should be added where necessary to aid the comprehension of words such as those naming parts of the body (Participants E and R) and to accommodate the needs of visual learners (like Participant M). The use of visual illustrations might also improve the SemiMed interface, making it more vivid and attractive (Participants L and R). Fourth, the inclusion of examples was highly recommended, as the majority of participants stressed the importance of seeing a word in context to better understand it and use it correctly. A sentence example would be "just fine" for this (Participant H). Furthermore, participants F, L and Q said that they spent a considerable amount of time reading longer texts (a paragraph or page of relevant documents) retrieved from a Google search. From their experiences it can be inferred that example sentences would both save time and satisfy the need to learn words from context.

5 OVERALL REFLECTIONS AND FUTURE PLANS

This study set out to pilot SemiMed, a new lexicographical resource of semi-technical medical vocabulary which is being developed in response to semantic deficiencies resulting from the reliance on word form frequency in the MWL. The development of SemiMed is based on theoretical premises of lexical semantics that have not been observed in current resources. The results of the pilot study show that SemiMed has some significant advantages over other resources, especially the radial visualizations of semantic relationships (polysemy and homography), which are the fruit of the consideration of lexical semantic theories during its development. SemiMed also addresses semantic deficiencies in the MWL because it takes into account word meaning frequency together with word form frequency. This methodological approach may pave the way for future studies which attempt to improve word form frequency-based wordlists. However, due to the limited timeframe of the study, and limited resources, SemiMed is not without flaws. The enhancement of its platform and provision of visuals, examples and pronunciation aspects are key areas that deserve further study. It is strongly believed that if these shortcomings are addressed, SemiMed has the potential to be of great benefit to EMP learners.

> This methodological approach may pave the way for future studies which attempt to improve word form frequency-based wordlists.

6 ACKNOWLEDGEMENTS

The endeavour to develop and pilot SemiMed could not have been possible without the A.S. Hornby Educational Trust, who financed this research project. I am grateful to the ASHDRA panel of experts and to my supervisors, Dr. Julia Miller and Dr. Stephen Kelly, for their valuable academic support from the early stage of SemiMed development to piloting SemiMed to HUMP participants. Thanks should also go to Ms. Maureen Goldfinch, my research assistant, and Mr. Phan Thanh Luan, IT support officer from HUMP, who designed electronic formats of SemiMed headwords and set up the Moodle platform. I would like to acknowledge Dr. Nguyen Thi Anh Phuong, Associate Head, and Ms. Hoang Thi Thanh Nhan, officer of Department of Science, Technology and International Relations, HUMP, who helped me liaise with the

participants, and all 18 HUMP medical students, who made themselves available after hours for Zoom meetings.

REFERENCES

Atkins, B. T. S. (1991). Building a lexicon: The contribution of lexicography. *International Journal of Lexicography*, *4*(3), 167-204.

Baker, M. (1988). Sub-technical vocabulary and the ESP teacher: An analysis of some rhetorical items in medical journal articles. *Reading in a Foreign Language*, *4*(2), 91-105.

Cantos, P., & Sanchez, A. (2001). Lexical constellations: What collocates fail to tell. *International Journal of Corpus Linguistics*, *6*(2), 199-228.

Cantos, P., Sanchez, A., & Almela, M. (2009). An attempt to formalize word sense disambiguation: Maximizing efficiency by minimizing computational cost. *RESLA*, 22, 77-88.

Chabner, D. E. (2020). *The language of medicine*. USA: Elsevier US.

Cohen, A., Glasman, H., Rosenbaum-Cohen, P. R., Ferrara, J., & Fine, J. (1988). Reading English for specialized purposes: Discourse analysis and the use of student informants. In P. Carrell, J. Devine, & D. E. Eskey (Eds.), *Interactive approaches to second language reading* (pp. 152-167). Cambridge, UK: Cambridge University Press.

Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics.* Cambridge, UK: Cambridge University Press.

Evans, V. (2005). The meaning of time: Polysemy, the lexicon and conceptual structure. *Journal of Linguistics*, *41*(1), 33-75.

Evans, V., & Green, M. (2006). *Cognitive linguistics: An introduction*. Edinburgh, UK: Edinburgh University Press.

Farrell, P. (1990). *Vocabulary in ESP: A lexical analysis of the English of electronics and a study of semi-technical vocabulary*. (CLCS Occasional Paper No. 25). Dublin, Ireland: Trinity College, Centre for Language and Communication Studies.

Fraser, S. (2009). Breaking down the divisions between general, academic and technical vocabulary: The establishment of a single, discipline-based word list for ESP learners. *Hiroshima Studies in Language and Language Education*, *12*, 151-167.

Fraser, S. (2012). Factors affecting the learnability of technical vocabulary: Finding from a specialized corpus. *Hiroshima Studies in Language and Language Education*, *15*, 123-142.

Geeraerts, D. (2006). *Words and other wonders: Papers on lexical and semantic topics*. Berlin: Mouton de Gruyter.

Higgins, J. J. (1966). Hard facts. *ELT Journal*, *21*(1), 55-60.

Hoey, M. (2012). *Lexical priming: A new theory of words and language*. London, UK: Routledge.

Hsu, W. (2013). Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list. *Language Teaching Research*, *17*(4), 454-484.

Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain and Language*, *81*(1-3), 205-223.

Lakoff, G. (1987). *Women, fire and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago.

Le, C. N. N., & Miller, J. (forthcoming). A core meaning-based analysis of English semi-technical vocabulary in the medical field.

Li, E. S.-L., & Pemberton, R. (1994). An investigation of students' knowledge of academic and subtechnical vocabulary. In J. Flowerdew & A. K. K. Tong (Eds.), *Entering text* (pp. 183-196). Hong Kong: The Hong Kong University of Science and Technology.

Moerdijk, F. (2003). The codification of semantic information. In P. Sterkenburg (Ed.), *A practical guide to lexicography* (pp. 273-296). Amsterdam, NL: John Benjamins Publishing.

Perez, M. (2013). *Identification and analysis of the specialized vocabulary of British Law Reports: A corpus-driven study of this legal genre at the core of common law legal systems.* (Doctoral thesis), University of Murcia, Spain. Retrieved from https://www.tdx.cat/handle/10803/128621?fbclid=IwAR29vxLl5QEpAcg19TKI6ExdP_9GJ92NunxxFLyTHcJYvz4aOKaKcyv2i10#page=1

Rizzo, C., & Sanchez, A. (2010). Building new meanings in technical English from the perspective of the lexical constellation model. *Ibérica*, *20*, 107-126.

Ruhl, C. (1989). *On monosemy: A study in linguistic semantics*. New York: State University of New York Press.

Ruhl, C. (2002). Data, comprehensiveness, monosemy. *Studies in Functional and Structural Linguistics*, 171-190.

Tyler, A., & Evans, V. (2003). Reconsidering prepositional polysemy networks: The case of over. *Trends in Linguistics Studies and Monographs*, *142*, 95-160.

Watson-Todd, R. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes*, *45*, 31-39.

APPENDICES

Medical Scenarios

Medical Scenario 1: Bowel
Specialist: We'll have to do tests, but I think you may have a **tumour**[1] of the colon.
Patient (scared): What does that mean?
Nurse: It means . . .
Specialist: What colour are your **stools**[2]?
Patient: My what?
Nurse: When you go to the toilet. Your . . . What colour is it?
Patient: Black.
Specialist: It might be nothing, but we need to do an operation. You may have a tumour. It might be **benign**[3]. That means . . . Or we may have to do a **radical**[4] operation. That means . . .
Patient: You think I've got cancer?
Nurse: Maybe. But we don't know until we do the operation.
Patient: Oh. What does my **colon**[5] do?
Nurse: It **absorbs**[6] water and moves the waste along so it can be passed out.
Specialist: But don't worry. There is a very good chance of removing any cancer if we find it quickly.

Medical Scenario 2: Eye
Specialist: Hello, X. Thank you for coming today. Do you know why you're here?
Patient: Yes. I can't see properly.
Specialist: Can you look at me? Now look at the nurse. I want to see your eye **reflexes**[1].
Patient: What do you mean?
Nurse: That means . . .
Specialist: And in your case you have a **cataract**[2].
Patient: What's that?
Nurse: It means . . .
Patient: But I wear glasses.
Specialist: Yes. You are short sighted. But a cataract is a **chronic**[3] eye condition.
Patient: You mean it's bad?
Nurse: Maybe. It means it . . . But we can **resolve**[4] the problem with a small operation.
Patient: Why have I got a cataract?
Specialist: There could be lots of reasons. You also have diabetes, so that **predisposes**[5] you to cataracts too.
Patient: Predisposes?
Specialist: [Explains] . . . Your eyes are also **secreting**[6] more mucus than is normal.
Patient: Secret?
Nurse: [Explains] . . .
Specialist: But we can treat that with eye drops. We'll arrange your cataract surgery next time we see you. It's not urgent.

Medical Scenario 3: Heart
Patient: What's wrong with me?
Specialist: You have **acute**[1] coronary syndrome.
Patient: Cute? Like kittens?
Nurse: No, 'acute'. That means . . .
Specialist: Have you had any **prior**[2] **cardiac**[3] problems?
Patient: Uh?
Specialist: [Explains] . . .
Nurse: Your heart is what makes your blood **circulate**[4].

Patient: Circle?
Nurse: [Explains] . . .
Specialist: We thought maybe there was a **shunt**[5] between the right and left sides of your heart. In other words, . . . But now we know that's not the case.
Patient: So I didn't have a heart attack?
Nurse: No. You had pain **radiating**[6] into your arms.
Specialist: [Explains] . . . But we are sure it's acute coronary syndrome. Make an appointment and I'll see you again next week.

Medical Scenario 4: Liver
Specialist: You have a problem with one of the **lobes**[1] in your **liver**[2].
Patient: Lobes? Like ear lobes?
Nurse: No. Your liver is divided . . .
Patient: What is my liver, anyway? What does it do?
Nurse: Your liver . . .
Specialist: You have a **disorder**[3] of the lobe. That means . . . There is a risk of infection. It's only **moderate**[4].
Patient: Moderate?
Nurse [Explains] . . .
Specialist: We'll need to cut through the **fascia**[5].
Patient (alarmed): Cut my face?
Specialist: No. 'Fascia' refers to . . . But don't worry. We'll **sedate**[6] you before the operation.
Nurse: [Explains] . . . You won't know anything about it till you wake up.
Patient (still a bit worried): OK.
Specialist: Don't worry. I've done this operation hundreds of times. You'll be fine.

Medical Scenario 5: Pregnancy
Nurse: Hello X. Please sit down. Do you know why you're here?
Pregnant patient: No.
Specialist: We would like to undertake some screening tests just to make sure your baby is growing well and doesn't have any **defects**[1]. A defect is . . . For example, a problem with the heart.
Nurse: We **conduct**[2] tests like this on all pregnant women. Mostly it's fine, but sometimes, towards the end of the pregnancy if the baby is not growing enough, we might have to **induce**[3] labour.
Pregnant patient: Induce?
Nurse: That means . . .
Specialist: Sometimes the problem is **compounded**[4] by a range of factors.
Pregnant patient: Arrange the factors?
Nurse: No. 'Compound' means . . . Many things could happen. For instance, some cells can **migrate**[5] from the baby to the mother and cause problems for the mother.
Pregnant patient: I'm not emigrating!
Specialist: No. It means . . . Some cells like carbon dioxide **diffuse**[6] from the baby to the mother which is normal. It allows the CO2 produced by the baby to cross over into the mother so she can breathe it out.
Nurse: Usually everything is fine. We'll make an appointment for a blood test and scan for you next week.
Pregnant patient: Thank you