Researching vocabulary depth in a multilingual Swedish elementary school

Researching vocabulary depth in a multilingual Swedish elementary school

Christine COX ERIKSSON
Dalarna University
SE-791 88 Falun
cce@du.se

Ziel dieses Beitrags ist es, Möglichkeiten und Herausforderungen bei der Untersuchung von Wortschatztiefe im Kontext mehrsprachiger Primarschulen aufzuzeigen. Im theoretischen Teil wird ein Überblick über die Messung von Wortschatztiefe anhand von Wortdefinitionen und Assoziationsaufgaben in der internationalen und schwedischen Forschung gegeben. Der empirische Teil beschreibt Datenerhebungen mit einem neu entwickelten Definitionstest und dem Kent-Rosanoff-Assoziationstest mit 92 Zweit- bzw. Fünftklässlern an einer schwedischen Schule mit hohem Anteil an Zweitsprachlern. Thematisiert wird die Bewertung der Definitionen von Schülerinnen und Schülern sowie damit zusammenhängende Herausforderungen. Die Ergebnisse, die nach Alter, Aufenthaltsdauer in Schweden und Sprachhintergrund analysiert wurden, werden im Hinblick auf die bestehende Forschung und methodische Überlegungen diskutiert. Angesprochen werden der Mangel an Forschung und verfügbaren Testinstrumenten in Bezug auf Wortdefinitionen sowie an geeigneten Worthäufigkeitslisten, die auf mündlicher Sprache in einem Grundschulkontext basieren. Das in diesem Artikel beschriebene Vorgehen und die Herausforderungen sowie die Ergebnisse stellen einen Beitrag zu unserem Wissen über die Wortschatztiefe bei mehrsprachigen Schülerinnen und Schülern dar und sind für Forscher und Pädagogen gleichermassen von Interesse.

Stichwörter:
Wortschatztiefe, Erstspracherwerb, Zweitspracherwerb, Wortdefinitionen, Wortassoziationen, Primarschule, Mehrsprachigkeit.

Keywords:
vocabulary depth, L1 acquisition, L2 acquisition, word definitions, word associations, elementary school, multilingualism.

1. Introduction

This article focuses on issues related to investigating vocabulary knowledge in Swedish, specifically lexical depth, in an elementary school where approximately 75% of pupils have a migration background and only 30% of parents have post-secondary education. Understanding how words are stored in the mental lexicon is a complex undertaking, even within the context of first language acquisition. Vocabulary depth, a measure of how well words are known, has alternatively been referred to as lexical quality (Anderson & Freebody 1981) or lexical organization (Meara 1996). According to Meara, the vocabularies of second language (L2) speakers may not be as well structured as those of first language (L1) speakers. Thus, organization may be a way to distinguish learners at different proficiency levels. Meara views organization as a characteristic of a learner's entire vocabulary, not of individual words. How vocabulary depth is measured, as well as empirical results, depend on how the construct is conceptualized (Schmitt 2014).

Receptive word knowledge is easier to measure and show, for example through picture recognition. Productive knowledge involves at least some mastery of a word's spoken or written form, as well as meaning and use (Nation 2013). Vocabulary production and grammatical ability in children approximately 1 to 4 years of age can be measured using the Communicative Development Inventories (CDI; SECDI in Sweden; e.g. Cox Eriksson 2014), which represent parents' records of their children's early language development. For older children, however, other tools must be used to describe vocabulary development in terms of lexical quality.

1.1 Aim

The present article aims to discuss methodological choices made and challenges experienced in assessing pupils' Swedish vocabulary depth in a multilingual elementary school. The research project from which these experiences are taken aimed to describe and attempt to explain the variation in vocabulary development in pupils attending a multilingual school and, together with the teachers, devise new classroom approaches that facilitate vocabulary learning, regardless of the pupils' language backgrounds. Participants include 92 pupils in grades 2 and 5 (mean ages 8.4 and 11.4 years), with at least 15 different first languages. Choices discussed in this article include deciding how to measure vocabulary depth and what instruments to use. In particular, the article will explore assessing vocabulary depth through a word definition task and a word association task.

2. Background

This section presents an overview of international research on investigating vocabulary depth in both L1 and L2 language contexts, with adults and younger learners.

2.1 Conceptualizing and measuring vocabulary depth

Nation's (2013) well-known conceptualization of what it means to know a word breaks down lexical knowledge into three main categories: *form* (spoken and written forms, as well as word parts), *meaning* (concepts, referents and associations) and *use* (grammatical functions, collocations and constraints on use). Each category includes receptive as well as productive aspects. Using most of Nation's terms, Figure 1 illustrates a learner's knowledge of a specific word after a certain number of exposures, with the various aspects being mastered at different rates. In studies with multilingual individuals, it is especially relevant to consider that there may be varied mastery of word knowledge.

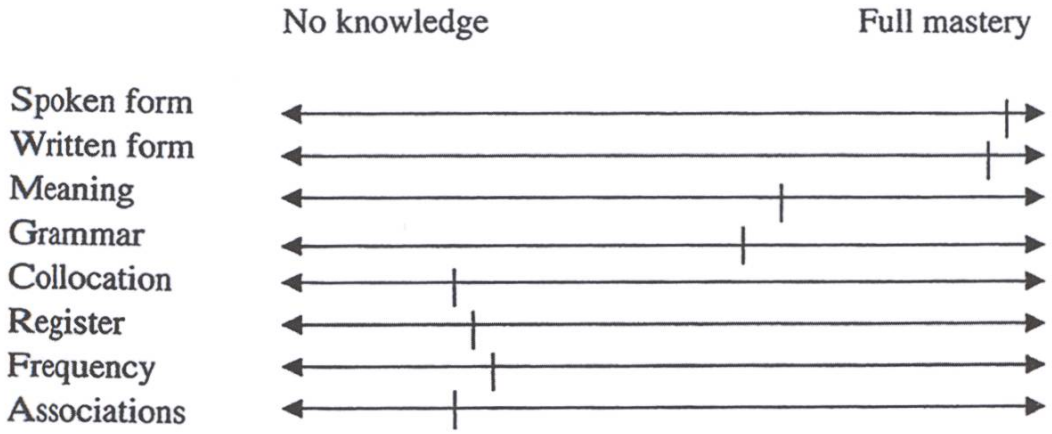

Figure 1: Developing knowledge of a word (Schmitt 2010: 38)

Meara and Wolter (2004: 87) have criticized the above conceptualization of vocabulary as components because it regards depth as knowledge of individual words, rather than a "test taker's entire vocabulary". Meara and Wolter compare two different ways of looking at vocabulary in Figure 2.

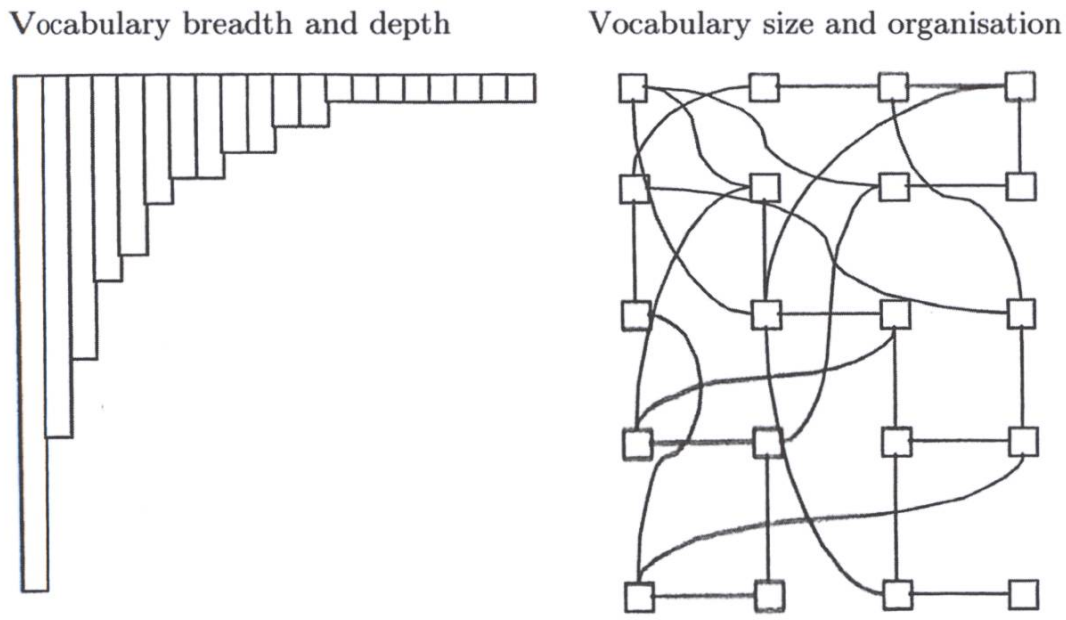

Figure 2: Two ways of looking at vocabulary (Meara & Wolter 2004: 89)

The image on the left shows that items learned earlier have more depth, as opposed to newly acquired words. However, this would imply that words are not related to one another. Therefore, Meara and Wolter provide the image on the right as a better model of vocabulary depth, illustrating how words are stored in

the mental lexicon as a network, with more links between individual words resulting in greater lexical organization.

Read (2004) outlines three different approaches that have been applied in research on L2 vocabulary acquisition. The first is *precision of meaning*, which implies the difference between a vague understanding of what a word means and more specific knowledge of that meaning. Secondly, there is *comprehensive word knowledge*, which involves many different aspects of a word, including orthographic, phonological, syntactical, collocational, and pragmatic attributes. This is comparable to Nation's component model described above. Thirdly, Read lists *network knowledge*, which means the incorporation of a word into the mental lexicon, where words are linked together in different ways. Read points out that any kind of vocabulary test will only reflect what a learner knows about the target words at the point the test is taken. Moreover, vocabulary tests often assess learners' declarative knowledge, which is descriptive and easier to verbalize, as opposed to procedural knowledge. The latter is more implicit and associated with word recognition or competent listening comprehension.

Precision of meaning has been exemplified by describing the progression in word learning in categories or stages. One classic scale is Dale's (1965: 898) four-stage description of how well a word is known, originally created to be used with elementary age L1 learners:

> Stage 1: "I never saw it before"; Stage 2: "I've heard of it, but I don't know what it means"; Stage 3: "I recognize it in context-it has something to do with …"; Stage 4: "I know it".

This is still a useful tool in developing awareness of word meanings among young learners.

2.2 Word definition tasks

Precision of knowledge can be investigated through definition tasks, often used in assessing children's L1 or L2 vocabulary knowledge. Benelli et al. (1988) conducted three studies to investigate the development of word definitions among children (5- and 7-year-olds) and adults. The task included defining nine common nouns that fall in three categories. Responses were classified on a scale ranging from perceptual appearance and functional definitions to different types of categorical definitions. The latter included either definitions with superordinates alone, those with some kind of specification added to the superordinate, or generic categorical definitions (i.e. 'thing' or 'object'). The youngest children gave the fewest definitions containing superordinates, the 7-year-olds used more, and the adults used most. A new group of adults judged the best definitions to be those including both categorical items and specific information about the target word. It was also found that by the time children are 10 years old, their definitions resemble adult models.

Kurland & Snow (1997) assessed the definitional skill of children in low-income families over time (from 5;3 to 10;10[1]) and compared children's definitions to those of their mothers. The children were asked to define eight nouns and answers were assessed on a 7-point scale for definitional quality, including various kinds of clauses and descriptors. Like Benelli et al., Kurland & Snow found that 9 to10-year-olds had also reached "adult levels of definitional skill", with approximately two thirds of the children giving better definitions than their low-income mothers. The authors conclude that "definitional skill is related to being part of an academic culture" (Kurland & Snow 1997: 603).

In a study aiming to investigate the role of oral vocabulary in reading skills (Ouellette 2006), 4th grade pupils were asked to define 32 words, including nouns, verbs and adjectives. Answers to the definition test were scored on a 0–3-point scale, based on the number of semantic features given. Other definition tasks for younger children (preK-1st grade) have been scored along a continuum, with fewer points for contextual responses and more for decontextualized responses (e.g. Coyne et al. 2009). More recently, Juska-Bacher et al. (2021) analyzed functional, descriptive and categorical word definitions in both children and adults on a scale from contextualized to decontextualized.

Two studies by Vermeer (2001) with L1 and L2 Dutch-speaking children illustrate different ways of conceptualizing how to assess vocabulary depth. The first tested breadth with a picture vocabulary test and a description task, where kindergarten children were asked to explain or describe the meaning of 27 words. Depth was measured through an association task where children provided formal definitions of 10 words, external features and component parts, functional and material characteristics, as well as use. Children's responses were described qualitatively using an association network for each word. L1 children received higher scores than L2 children, who were able to provide characteristics, but lacked the verbal ability to describe them. The second study included a picture vocabulary test (receptive breadth) and a task asking for descriptions, characteristics or definitions of words. Results indicated that the input frequency of the words strongly correlated with the probability of knowing a word, especially in the description task. Vermeer's operationalization of vocabulary breadth and depth with young children exemplifies difficulties in comparing results across studies (Schmitt 2014).

In other research from the Netherlands, Verhallen & Schoonen (1993, 1998) compared lexical knowledge in L1 and L2 elementary school pupils using an extended word definition task. In a structured interview situation, bilingual Turkish-Dutch 3rd and 5th graders all born in the Netherlands were asked to give

[1] Within the field of child language research, age is often denoted for example as 5;3 (five years and three months).

as many meaning aspects as possible to six stimulus words. Results indicated that the pupils were better able to explain the words in Dutch than in Turkish. Comparisons with L1 pupils showed that the L2 pupils gave fewer meaning aspects as well as less well-developed answers, which included fewer paradigmatic meaning aspects. The relationship between paradigmatic (same-class, as in *sit-stand*) and syntagmatic (different class, *sit-chair*) word knowledge was similarly investigated in an American study (Ordóñez et al. 2002), where bilingual (Spanish-English) 4th and 5th graders provided superordinates, definitions and rich object descriptions of six familiar concrete nouns. Results indicated that pupils' ability to produce paradigmatic responses in one language often transferred to the other.

2.3 Word association tasks

The classic Kent-Rosanoff word association test (Kent & Rosanoff 1910) assesses the ability to make hierarchical associations, essentially measuring network knowledge (Read 2004). The original list is comprised of 100 nouns and adjectives and has been used in studies with both children and adults. The test involves presenting participants with a single stimulus word, such as *cat*, and asking them to give the first word they associate with that word. Scholars have systematized responses into categories, including clang associations, based on phonological similarity to the stimulus word (*hat-cat*), syntagmatic associations, based on semantic principles (*cat-meows*), and paradigmatic associations (*cat-animal*), often denoting the hierarchical category in which the stimulus word belongs. The so-called syntagmatic-paradigmatic shift has been found to occur at some point between 6 and 10 years of age and is an indication of increased lexical knowledge with the ability to form paradigmatic associations (Nelson 1977). According to Nelson (1977), the shift may not only indicate conceptual reorganization, but also task capacity. One study with young first and second language learners has indicated that the shift to more paradigmatic associations may occur later in multilingual children (Schoonen & Verhallen 2008).

Read (1993) created the *Word Associates Format* (WAF), where responses to a stimulus word are selected, rather than supplied. This was thought to be a more practical option for L2 university learners than the classic Kent-Rosanoff test. In this test, a target word (for example *contract* is followed by six or eight other words, which may or may not be related to the stimulus word (i.e., *agreement, confident, formal, notice, sign, special*). Answers can be scored with respect to the type of relationship between the target word and associated words, e.g., *agreement* (paradigmatic; synonym), *sign* (syntagmatic; collocate) and *formal* (analytic; word which represents a key element in the meaning of the target word). Research has generally shown that scores on tests using WAF are predicative of skills such as reading comprehension (Fitzpatrick & Thwaite 2020). Schoonen & Verhallen (2008) adapted Read's WAF for use with upper elementary pupils (grades 3-6) and developed the Word Association Task

(WAT). Pupils choose three words out of six that best go together with the stimulus word and draw connecting lines (see Figure 3 for correct answers). *Monkey* is a semantically less related word and therefore, a less decontextualized answer.

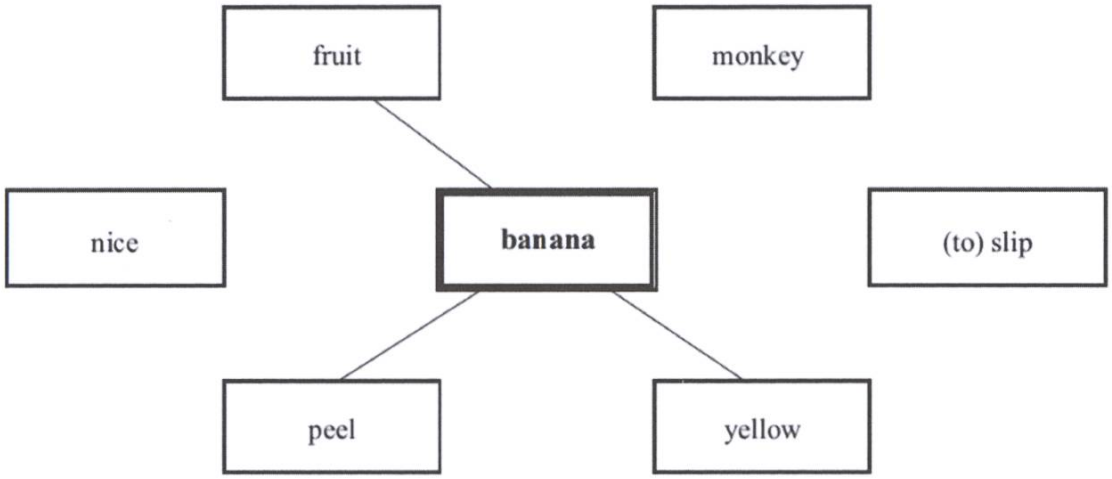

Figure 3: Sample item from the WAT (translated from Dutch; Schoonen & Verhallen 2008: 219)

In a recent review article, Fitzpatrick & Thwaite (2020) highlight several methodological challenges involved in comparing results of word association research, including specific definitions of the paradigmatic/syntagmatic/clang categories, as well as the number of cue words used and elicited responses per cue. They also caution against using word association results to accurately reflect the organization of the mental lexicon, as no single method can be used to do this.

3. Research in the Swedish context

Many Swedish studies on children's vocabulary development are from the field of speech therapy. Andersson et al. (2019) investigated the role of multilingualism and socioeconomic status in Swedish L1 and L2 first and second graders' results on the Clinical Evaluation of Language Fundamentals (CELF4; Semel et al. 2013). A composite of social deprivation measures was found to account for a much greater proportion of variance than the children's multilingualism. As one of few Swedish studies on word definitions, Erlandsson & Yhlen (2019) investigated Swedish L1 and L2 8-year-olds' definitions of 10 words common in a school context. They followed the scoring principle (0-3 points) used by McGregor et al. (2012): 0 points for no correct information; 1 point for some meaningful relationship to the target word; 2 points for minimal definitions; 3 points for more than minimal accurate information. Erlandsson & Yhlen found that while L1 pupils' definitions gave more information than their L2 peers, participants rarely received full scores for their definitions. There is a fair

amount of research on grammatical development in Sweden, for example work by Gisela Håkansson. She has recently been involved in a project using the new assessment tool Cross-linguistic Lexical Tasks (LITMUS-CLT; Haman et al. 2017), which explores young learners' productive and receptive knowledge of nouns and verbs in multiple European languages. A recent doctoral thesis investigating oral narrative competence in L1 and L2 Swedish 4-6-year-olds (Lindgren 2018) included vocabulary production scores using CLTs.

Namei (2002), who translated the original 100-word Kent-Rosanoff list into Persian and Swedish, compared the L2 and L1 mental lexicon in terms of syntagmatic-paradigmatic development and similar/dissimilar organization in 100 bilinguals aged 6 to 22 and 50 L1 speakers of both languages. Clang responses were most common with unknown or less frequent words, syntagmatic responses were given when words were partially known, and paradigmatic responses were given when words were highly frequent or well-integrated in the lexicon. Mikoczy & Nyman (2008) used Namei's Swedish translation with bilingual Swedish-Arabic 4th graders, calling for a shorter version of the list. Johansson & Wahlstrand (2010) shortened the list to 50 words and provided an extended assessment guide. The shortened version, also translated into Arabic, has been used in many studies, including one comparing lexical development in bilingual 4th graders who had received bilingual instruction as opposed to those who only received instruction in Swedish (Salameh 2011). A significantly higher proportion of pupils receiving bilingual instruction showed hierarchical lexical organization in both languages, but this was not correlated to lexical size. Nilsson & Svenbe (2017) used the 50-word Swedish list, comparing lexical organization and size in three groups of teenagers and young adults. Results indicated that paradigmatic responses dominated in all groups, followed by syntagmatic responses.

Other Swedish studies have used the Kent-Rosanoff test to compare lexical organization in monolingual and bilingual children with and without developmental language impairment. Holmström et al. (2016b) used conceptual scoring in the word association task for bilinguals, where knowledge of words in both languages was combined. Bilingual children with language impairment received significantly higher paradigmatic conceptual scores compared to scores based on one of their languages alone. Their scores were also higher than those of monolingual children with language impairment. However, when both productive and receptive measures were included, smaller vocabulary sizes were evident in the bilinguals with language impairment as opposed to typically developing bilinguals (Holmström 2015). Holmström et al. (2016a) point out that the Kent-Rosanoff list was not originally constructed to study lexical organization, calling for further development of a word association list considering word frequency, part of speech and mean age of acquisition of words. A more recent study (Sandgren et al. 2020) also aimed to investigate vocabulary breadth, depth, and fluency, and used a semantic depth score in

comparing word associations in children aged 6 to 9 with developmental language disorder (DLD) to those with typical development. To calculate the semantic depth score, paradigmatic (2 points each) and syntagmatic (1 point) associations were summed. The typically developing group had significantly higher semantic depth scores than the DLD group.

In summary, this literature review has shown that vocabulary depth has been conceptualized and operationalized in different ways. This, along with great variety in the scoring of definition and association tasks, makes comparisons across studies a challenge. However, the variety of tools available do allow researchers to tailor assessment instruments to a specific student population. As the current study included pupils with varying reading ability in Swedish, we wanted to make sure that we were assessing vocabulary knowledge and not reading ability. Thus, we chose two test instruments that were administered orally, even if the WAT has already been used internationally with young pupils.

4. Measuring depth in the current study

Data for the current project was collected from November 2018 to early April 2019 at a linguistically diverse elementary school in central Sweden. Participants were 92 pupils (47 girls; 45 boys) in grades 2 (n = 54; mean age 8.4 years) and 5 (n = 38; mean age 11.4 years). Pupils were diverse in terms of how long they had lived in Sweden as well as whether they had Swedish as L1 or L2 (see Table 1).

	Years in Sweden				Total
Grade	1–4 (SL2)	5–7 (SL2)	Born in Sweden (SL2)	Born in Sweden (SL1)	
2	14	11	18	11	54
5	7	8	13	10	38
Total	**21**	**19**	**31**	**21**	**92**

Table 1: Years in Sweden and language background for all participants (N = 92) in grades 2 and 5.

4.1 Word definitions

Due to the lack of available normed instruments for measuring vocabulary depth in Swedish, a newly constructed word definition task (Edquist 2021) was developed for use in the current project. Following Edquist's previous trials, items include words of varying frequency and difficulty level, as well as a mix of everyday and academic words. The current project used a list of 12 items comprising four verbs (*applådera* [applaud], *inbilla sig* [imagine], *fantisera* [fantasize] and *prioritera* [prioritize]), five nouns (*lön* [salary], *ved* [firewood], *fe*

[fairy], *kapitel* [chapter] and *dygn* [one day and night/24 hours]) and three adjectives (*generös* [generous], *fridlyst* [protected, as in a protected species] and *stolt* [proud]).

The definition task attempted to assess pupils' ability to spontaneously define words with only a limited amount of prompting. After the original question *What does xxx mean?*, additional prompting was limited to questions such as: *Can you say more/add anything/explain?* or *What do you mean?* The pupils were told to imagine that the researcher came from another planet and did not know the meaning of many words. The goal of this test was to obtain some measure of depth as "precision of meaning" (Read 2004) or the "extent of semantic representation" (Ouellette 2006). Inspiration was also taken from categorizations of definitions outlined in several previous studies (Benelli et al. 1988; Kurland & Snow 1997). Like Ouellette (2006), our study included words from different parts of speech (although fewer words) and used a 3-point scoring scale, in contrast to studies using only concrete nouns and employing different scoring scales (Benelli et al. 1988; Kurland & Snow 1997; Ordóñez et al. 2002). Our 3-point scale (see below) is somewhat more specific than that used by McGregor et al. (2012) and Erlandsson & Yhlen (2019).

4.1.1 Scoring

Pupils' answers were scored on a scale of 0-3 points, depending on how much and what kind of information was provided. A new scoring rubric, which was continually revised, varied slightly for the different parts of speech. For example, for the noun *fe* [fairy], one point was awarded for a description (*a fairy is a fairy tale character*), one point for an example (*the tooth fairy)*, and one point for a characteristic (*a fairy can fly*).

To further illustrate scoring, answers (author's translation) are shown for a verb (*applådera*), a noun (*ved*) and an adjective (*generös*):

> *Applådera* (3 points): *When you dance or sing on the stage* (example), *if people think it is good* (purpose - why) *they do it, they clap their hands* (description).
> (2nd grader, Tigrinya speaker, 2 years in Sweden)
> *Ved (2 points): It's wood (description), thick, that you can heat with (purpose).*
> *(5th grader, Kurmanji speaker, born in Sweden)*
> *Generös* (0 points): *If someone is good-looking, or? Generous is like if someone is nice, courageous and intelligent, and ehm what else, smart, funny*.
> (2nd grader, Kurmanji speaker, born in Sweden)

One of the many challenges involved in the scoring process regarded how to systematize pupils' answers in the rubric and decide what constituted a full 3-point answer. For *dygn* [one day and night], one point each was awarded for using the word in an example (*it's like around the clock*) or providing a description (*24 hours*). In the end, we decided that providing both *24 hours* and *one day and night* constituted a full understanding of the word and was worth 3 points, even if this could only be considered two elements. We went through a similar process with the word *lön* [salary]. After deciding to award 1 point for a description (*it's money*), 1 point for an example (*maybe you work as a teacher*)

and 1 point for the purpose, or reason why (*when you work/from your job*), we decided that a relatively concise response (*you get money because you work*) could also be considered a 3-point answer. Wordy answers were not always worth more points. However, some talkative 5th graders gave very thorough answers with more examples or characteristics than were needed for the 3-point criteria. Other less talkative pupils (in both grades) gave shorter answers, even when the researcher asked follow-up questions. Individual personality traits may thus affect the answers given by participants, as well as factors such as language background, proficiency level, years in Sweden, and reasonably, socioeconomic status.

4.1.2 Results: Word definitions

Figure 4 shows the mean number of points awarded for the 12 items in the definition test, with information on frequency and difficulty levels in the legend (see below). Scores for 2nd graders and 5th graders are shown separately. Although there was great individual variation among the groups, 5th graders' definition scores were significantly higher (p = .000; two-tailed; t = -7.941) than those of the 2nd graders.

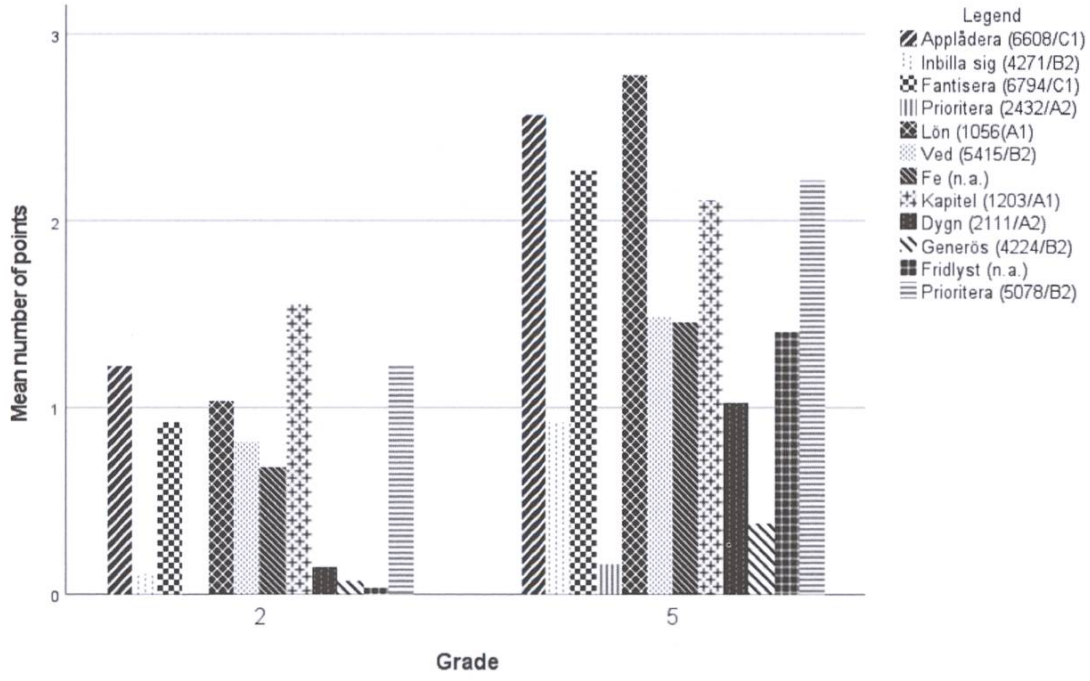

Figure 4: Mean number of points awarded for the 12 items in the definition test for all pupils grouped according to grade level. Frequency statistics and difficulty level are given for each word (KELLY/CEFR; the Swedish KELLY list and the Council of Europe's difficulty list are explained below).

As previous research shows that children's definitional ability approaches adult levels by approximately age 10 (Benelli et al. 1988; Kurland & Snow 1997), this is a reasonable result. In our study, the same words were easiest to define for both groups (*applådera, fantisera, lön, kapitel* and *stolt*), although to varying

degrees for younger and older pupils. The words that were most difficult to define for both ages were *inbilla sig, prioritera, generös, dygn* and *fridlyst*. None of the 2nd graders, and only 11% of the 5th graders could define *prioritera*. *Generös* was difficult for both 2nd and 5th graders (with 6% and 24%, respectively, receiving points at all). Only one 2nd grader could give a partial definition for *fridlyst*, while almost half of the 5th graders received a full score (16 of 37). Of the 19 5th graders who could not define the word, 17 were L2 Swedish speakers.

In the evaluation of results, comparisons were made with the Swedish frequency-based vocabulary list, KELLY (KEywords for Language Learning for Young and adults alike; Kilgarriff et al. 2014), which is based on a web-acquired corpus (SweWAC [2]) of 114 million words and incorporates the Common European Framework of Reference for Languages (CEFR; Council of Europe 2001) levels of difficulty. Frequency statistics for both the Swedish KELLY and CEFR levels are based on web texts aimed at L1 speakers. Despite efforts to create frequency-based word lists for Swedish L2 speakers, none of these are appropriate for research with children. A few comparisons between word frequency and our results illustrate the need for suitable lists. For example, the word *applådera* ranks 6608 in difficulty on the KELLY list and is a C1 word according to CEFR. However, it is commonly used in oral language, as well as in a school context. Therefore, it was one of the easier words for the pupils to define. *Prioritera*, on the other hand, ranks 2432 on the KELLY list and is an A2 word in accordance with CEFR. However, none of the second graders and only three of the fifth graders could provide any aspect of a definition. The verbs *inbilla sig* and *fantisera* are semantically related, but more different in Swedish than what they may seem to be in English. *Fantisera* is ranked 6794 in the KELLY list and is a C1 word according to CEFR. However, it is more common in a school context, especially regarding story writing. Pupils are encouraged to think up stories and use their imagination, which is *fantasi* in Swedish. *Inbilla sig*, on the other hand, is ranked somewhat more frequent and easier than *fantasize* (4271 KELLY /B2 CEFR). Nevertheless, the meaning is less related to children's everyday classroom experience, as it refers to something imagined in the sense of "to delude oneself". Our results underscore the need for more appropriate frequency lists, as previous research indicates that input frequency correlates with children's word knowledge (e.g., Vermeer 2001), but also that word frequency itself may vary in different contexts (Nation 2013; Schmitt 2014).

Figure 5 contrasts definition results for Swedish L1 speakers with those for pupils with Swedish as L2. Group means are higher at each grade level for Swedish L1 pupils (14.9 compared to 6.1 for grade 2; 24.3 compared to 16.7 for grade 5). In grade 2, the easiest words for L2 pupils to define were *kapitel* and

[2] The Swedish Web Acquired Corpus (SweWAC) was the main corpus, consisting of texts from the internet, which was used in the construction of the Swedish KELLY list (Kilgariff et al. 2014).

stolt, the only two words with means greater than 1 point. Apart from *prioritera*, Swedish L1 pupils in grade 2 received more points than L2 pupils for all words.

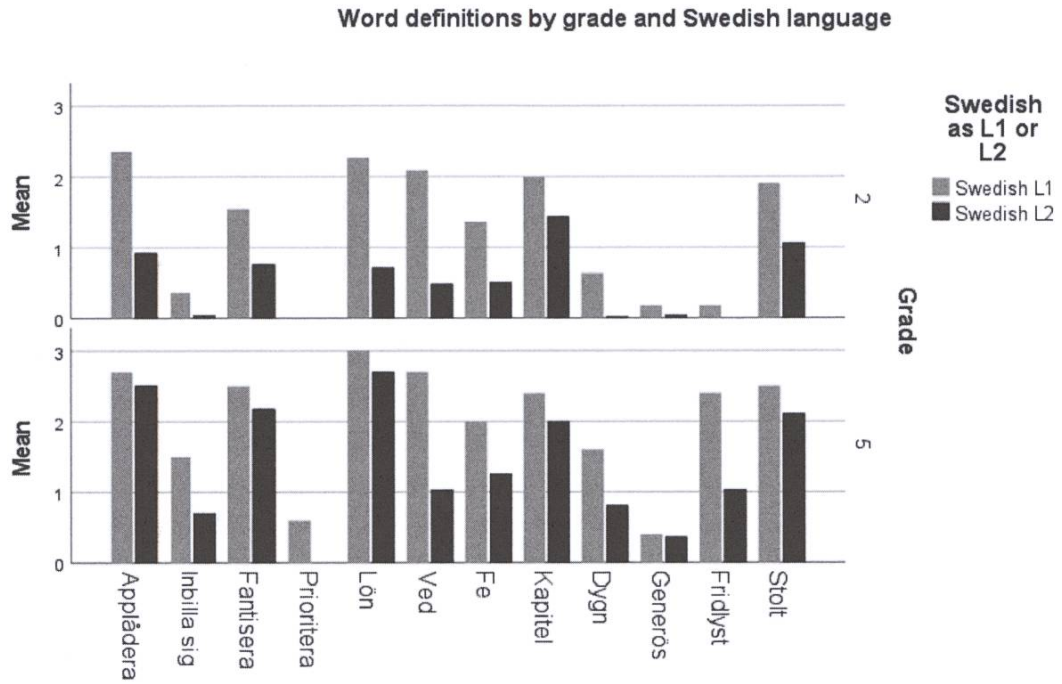

Figure 5: Mean number of word definition points according to grade and Swedish L1/L2

In grade 5 the Swedish L1 pupils received on average more than 2 points for eight of the 12 items, whereas those with other language backgrounds only had means above 2 points for five words: *applådera, fantisera, lön, kapitel* and *stolt*. Words such as *ved* and *fe* may have been culturally very different for the pupils. Although frequency statistics are missing for the word *fridlyst,* it represents a subject-specific word and was much more difficult for Swedish L2 5th graders.

4.2 Kent-Rosanoff word association test

In our study, we used the 50-word Kent-Rosanoff (K-R) list in Swedish only as a measure of vocabulary depth. Responses were coded as clang, syntagmatic, paradigmatic, other or no answer. Two groups of pupils were tested by a student, and three groups were tested by the present author. Administration and coding of the test was thoroughly discussed with the student together with a project colleague. As part of the student's essay (Norman 2019) inter-rater reliability was calculated between three independent scorers of the 50 responses from six pupils. As we found our score of 87% to be less than adequate, lengthy discussions involving scoring guides from two different sources (Nilsson & Svenbe 2017; Språkens hus, n.d.) were conducted to obtain more uniform coding.

4.2.1 Scoring difficulties

Assessment of pupils' answers was very time-consuming. Despite instructions to give one word in response to the stimulus word, many pupils answered with longer phrases instead, which could be classified as paradigmatic, syntagmatic or 'other', depending on the type of phrase and the part of speech of the headword: different part of speech for syntagmatic responses (*doctor – sometimes I'm afraid of shots)*; same part of speech for paradigmatic responses (*doctor – someone who gives medicine*). Multiword responses without semantic connection to the stimulus word were classified as 'other' (*woman – one who stands somewhere*). In general, as noted by Fitzpatrick & Thwaite (2020), scoring difficulties often regarded choices between syntagmatic and paradigmatic responses, or whether a word had a strong enough semantic relation to the stimulus word.

4.2.2 Results for Kent-Rosanoff test

Figure 6 presents results for all 92 pupils on the Kent-Rosanoff test, grouped according to grade level and time in Sweden. Pupils in grade 2 with language backgrounds other than Swedish used a larger percentage of clang responses compared to the L1 pupils, while pupils in grade 5 used proportionately very few or none overall, showing both age and proficiency effects (see Namei 2002).

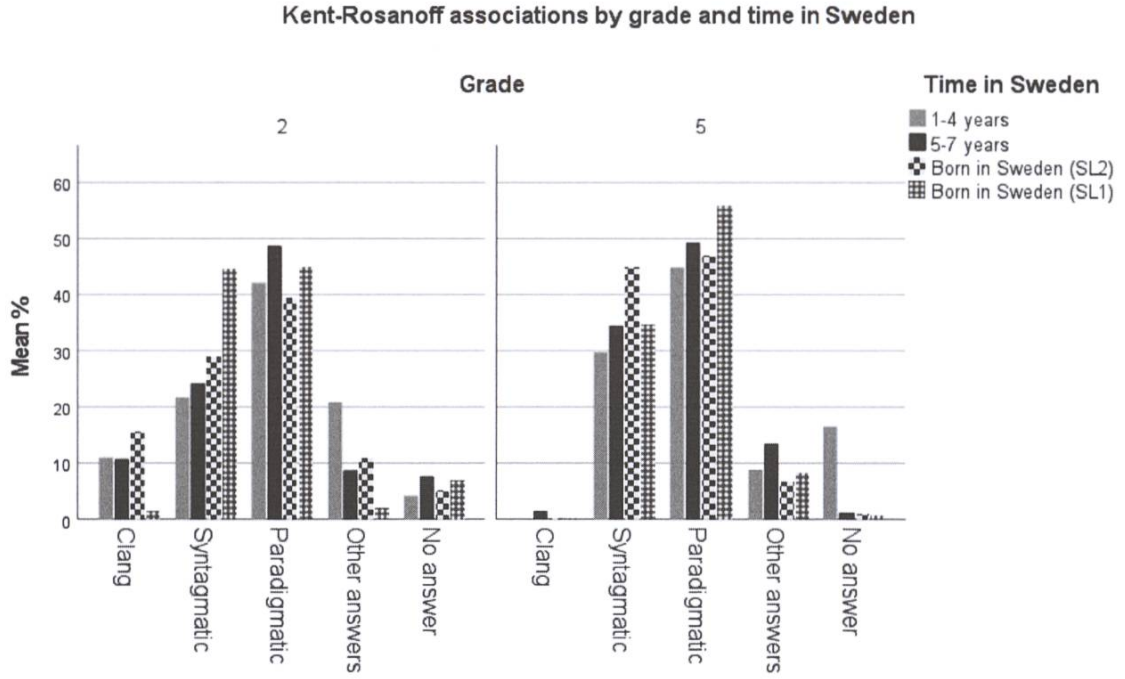

Figure 6: Proportions of clang, syntagmatic, paradigmatic, other and no answers for all pupils in grades 2 and 5 (N = 92), grouped according to time in Sweden. For numbers of pupils in each group, see Table 1.

Swedish L1 pupils in grade 2 responded with equal percentages of paradigmatic and syntagmatic responses, while L1 pupils in grade 5 used considerably more paradigmatic responses, which may be evidence for the syntagmatic/ paradigmatic shift. Grade 5 pupils with shorter length of residence in Sweden also gave more paradigmatic responses, but the trend was not as clear for the L2 pupils born in Sweden, indicating that the shift may occur later in these multilingual children (Schoonen & Verhallen 2008). The percentages of responses classified as 'other' or no answers at all can be due to both age and proficiency level.

Following Sandgren et al. (2020) we calculated a semantic depth score for each pupil based on the number of syntagmatic (1 point) and paradigmatic (2 points) associations. As a group, 5th graders had significantly higher semantic depth scores than the 2nd graders (p = .011; t = -2.586), which may also indicate a higher level of lexical organization in their vocabulary.

5. Discussion

The aim of this article has been to discuss choices made and challenges experienced in assessing pupils' Swedish vocabulary depth in a multilingual elementary school. As highlighted by Schmitt (2014), empirical results depend first on how depth is conceptualized, and then on how it is measured. We see word definitions and associations as complementary measures of vocabulary depth, illustrating different aspects of pupils' vocabulary knowledge. Our study presents results using a new definition test, providing valuable new knowledge, considering the paucity of appropriately normed tests in Sweden. We chose to use a definition test with indirect prompting, where pupils were told to imagine that the researcher came from another planet. This contrasts with other methods such as Verhallen & Schoonen's (1993, 1998) extended interview model. We found that older pupils and those with Swedish as L1 had higher definition scores at group level, while L2 pupils often lacked the verbal ability to describe words (see Verhallen & Schoonen 1998; Vermeer 2001). These results are perhaps not surprising, as age and language proficiency are important factors influencing vocabulary development. However, many individual pupils in grade 2, including speakers of Swedish as L1 and L2, received higher scores than older pupils.

We also chose to use the Kent-Rosanoff association test, which had previously been used in Sweden. Our results seemed to indicate differences in the organization of L2 learners' vocabularies compared to L1 learners (Meara 1996; Namei 2002; Schoonen & Verhallen 2008) as well as between younger and older pupils, although this may also be related to task capacity (Nelson 1977). In addition, critics question the use of any one task to reflect lexical organization and call for more large-scale research focused on L2 word association networks (Fitzpatrick & Thwaites 2020). Moreover, our two assessments give an

indication of pupils' lexical knowledge based on the words included, and capture what pupils know at that specific time point (Read 2004). Choosing to use a different association task would likely have produced different results. If we had tested pupils' L1 as well and used conceptual scoring on the Kent-Rosanoff test as did Holmström et al. (2016b), we would have obtained a better picture of the multilingual pupils' lexical knowledge. Due to the large number of first languages among the pupils and our limited time frame, this was not possible.

Participants were not only linguistically diverse but had varying lengths of residency in Sweden. This contributed to the great overall variation in our results. We lacked specific information regarding pupils' previous educational experiences and exposure to Swedish which confounds the categorization of all multilingual pupils born in Sweden as L2 learners, even though most of them had two foreign-born parents. It is likely that some of these pupils may have two first languages. In Sweden today, where input in English is pervasive, very few young pupils, including those with Swedish language background, can be considered strictly monolingual. We also lacked background information regarding parental education and other socioeconomic indicators, which have been shown to account for a large degree of variation in pupils' vocabulary knowledge (Andersson et al. 2019). These factors, as well as the use of a newly constructed definition test without a set scoring rubric, influence the reliability of our results.

We had originally planned to carry out our battery of tests, which also included a newly constructed naming test and a test of academic word knowledge, at two time points. This would have given us a picture of development over time. However, our schedule was disrupted by the Corona pandemic. Still, we were able to provide a snapshot of what pupils in a linguistically diverse elementary school know or can articulate. Our results also demonstrate great individual variability in pupils' vocabulary knowledge, which is to be expected among multilingual learners, but also in young learners' language development overall.

Despite uncertainties in our methods, our qualitative and quantitative analyses represent valuable knowledge regarding vocabulary development among multilingual pupils and are of interest for researchers and educators alike. We hope that our study may inspire the development of methods to investigate vocabulary depth, as well as the creation of frequency lists appropriate for pupils in an elementary school context. This would deepen our understanding of age, frequency, and proficiency effects on word learning.

REFERENCES

Anderson, R. C. & Freebody, P. (1981). Vocabulary knowledge. I: 1. In T. Guthrie (ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newark, DE: International Reading Association.

Andersson, K., Hansson, K., Rosqvist, I., Lyberg Åhlander, V, Sahlén, B. & Sandgren, O. (2019). The contribution of bilingualism, parental education, and school characteristics to performance on the clinical evaluation of language fundamentals: Swedish. *Frontiers in Psychology*, *10*, 1586.

Benelli, B., Arcuri, L. & Marchesini, G. (1988). Cognitive and linguistic factors in the development of word definitions. *Journal of Child Language*, *15(3)*, 619-635.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment.* Press Syndicate of the University of Cambridge.

Cox Eriksson, C. (2014). *Children's vocabulary development: The role of parental input, vocabulary composition and early communicative skills* (Doctoral dissertation, Department of Special Education, Stockholm University).

Coyne, M.D., McCoach, D.B., Loftus S., Zipoli, R. & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *The Elementary School Journal 110(1)*, 1-18.

Dale, E. (1965). Vocabulary measurement: Techniques and major findings. *Elementary English*, *42,* 895902, 948.

Edquist, G. (2021). *Expressivt ordförråd hos barn med lätt till måttligt svår hörselnedsättning.* [Masteruppsats]. Institutionen för neurovetenskap, logopedi, Uppsala universitet.

Erlandsson, J. & Yhlen, T. (2019). *"Skillnad betyder att det kan vara en skillnad mellan en person och en annan person" En jämförelse mellan en-och flerspråkiga skolbarns definitioner av skolrelaterade ord.* Lund, Sweden: Lund University.

Fitzpatrick, T., & Thwaites, P. (2020). Word association research and the L2 lexicon. *Language Teaching*, *53(3)*, 237-274.

Haman, E., Łuniewska, M., Hansen, P. Gram Simonsen, H., Chiat, S., Bjekić, J., Blažienė, A., Chyl, K., Dabašinskienė, I., Engel de Abreu, P., Gagarina, N., Gavarró, A., Håkansson, G., Harel, E., Holm, E., Kapalková, S., Kunnari, S., Levorato, C. Lindgren, J., Mieszkowska, K., Montes Salarich, L., Potgieter, A., Ribu, I., Ringblom, N., Rinker, T., Roch, M., Slančová, D., Southwood, F., Tedeschi, R., Müge Tuncer, A., Ünal-Logacev, Ö., Vuksanović, J. & Armon-Lotem, S. (2017). Noun and verb knowledge in monolingual preschool children across 17 languages: Data from Cross-linguistic Lexical Tasks (LITMUS-CLT). *Clinical Linguistics & Phonetics*, *31(11-12)*, 818-843.

Holmström, K. (2015). *Lexikal organisation hos en-och flerspråkiga skolbarn med språkstörning.* (Doctoral dissertation, Lund University).

Holmström, K., Salameh, E.K. Nettelbladt, U. & Dahlgren-Sandberg, A. (2016a). A descriptive study of lexical organisation in bilingual children with language impairment: Developmental changes. *International Journal of Speech-language Pathology*, *18(2)*, 178-189.

Holmström, K., Salameh, E.K., Nettelbladt, U. & Dahlgren-Sandberg, A. (2016b). Conceptual scoring of lexical organization in bilingual children with language impairment. *Communication Disorders Quarterly*, *38(1)*, 24-34.

Johansson, S. & Wahlstrand, E. (2010). *Lexikal organisation hos barn i förskoleklass och tredje klass. En omarbetning av Kent och Rosanoffs associationstest.* Lund, Sweden: Lund University.

Juska-Bacher, B., Brugger, L., Korthus, R. & Zangger, C. (2021). Definitionskompetenzen von Erst- und Zweitklässlern. Mit einem Ausblick auf die weitere Entwicklung. *Muttersprache 2,* 117-137.

Kilgarriff, A., Charalabopoulou, F., Gavrilidou, M., Bondi Johannessen, J., Kahklil, S., Johansson Kokkinakis, S., Lew, R., Sharoff, S., Vadlapudi, R, & Voldina, E. (2014). Corpus-based vocabulary lists for language learners for nine languages. *Language Resources and Evaluation*, *48*, 121-163.

Kent, G. H. & Rosanoff, A.J. (1910). A study of association in insanity. *American Journal of Psychiatry, 67(1)*, 37-96.

Kurland, B.F. & Snow, C.E. (1997). Longitudinal measurement of growth in definitional skill. *Journal of child language, 24(3)*, 603-625.

Lindgren, J. (2018). *Developing narrative competence: Swedish, Swedish-German and Swedish-Turkish children aged 4-6* (Doctoral dissertation, Acta Universitatis Upsaliensis).

McGregor, K.K., Berns, A.J., Owen, A.J., Michels, S. A., Duff, D., Bahnsen, A. J. & Lloyd, M. (2012). Associations between syntax and the lexicon among children with or without ASD and language impairment. *Journal of autism and developmental disorders, 42(1)*, 35-47.

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer & J. Williams (eds.), *Performance and competence in second language acquisition* (pp. 35-53). Cambridge: Cambridge University Press.

Meara, P. & Wolter, B. (2004). V_LINKS: Beyond vocabulary depth. In D. Albrechtsen, K. Haastrup, & B. Henriksen (eds.), *Angles on the English speaking world 4* (pp. 85-96). Copenhagen, Denmark: Museum Tusculanum Press.

Mikoczy, T. & Nyman, A. (2008). *Lexikal förmåga hos flerspråkiga svensk-arabiska barn i skolår 4. En undersökning med PPVT-III och Kent & Rosanoffs associationstest.* Lund, Sweden: Lund University.

Namei, S. (2002). *The bilingual lexicon from a developmental perspective: A word association study of Persian-Swedish bilinguals* (Doctoral dissertation, Stockholms universitet, Centrum för tvåspråkighetsforskning).

Nation, I.S.P. (2013). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Nelson, K. (1977). The syntagmatic-paradigmatic shift revisited: A review of research and theory. *Psychological bulletin, 84(1)*, 93.

Nilsson, A. & Svenbe, J. (2017) *Lexikal organisation hos ungdomar och unga vuxna.* Magisteruppsats. Linköpings universitet.

Norman, J. (2019). *Ordförråd hos elever i en mångkulturell skola: Lexikal organisation hos två klasser i årskurs 2 och 5.* Falun, Sweden. Högskolan Dalarna.

Ordóñez, C.L., Carlo, M.S., Snow, C.E. & McLaughlin, B. (2002) Depth and breadth of vocabulary in two languages: Which vocabulary skills transfer? *Journal of Educational Psychology 94(4)*, 719-728.

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98(3)*, 554.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10*, 355-371.

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 209-227). Amsterdam: John Benjamins.

Salameh, E.K. (2011). Lexikal utveckling på svenska och arabiska vid tvåspråkig undervisning. *Educare, 3*, 205-227.

Sandgren, O., Salameh, E.K., Nettelbladt, U., Dahlgren-Sandberg, A. & Andersson, K. (2020). Using a word association task to investigate semantic depth in Swedish-speaking children with developmental language disorder. *Logopedics Phoniatrics Vocology, 46(3)*, 1-7.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual.* Houndmills, Basingstoke, Hampshire, New York, NY: Palgrave Macmillan.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning, 64(4)*, 913-951.

Schoonen, R. & Verhallen, M. (2008). The assessment of deep word knowledge in young first and second language learners. *Language Testing*, *25(2)*, 211-236.

Semel, E., Wiig, E.H. & Secord, W.A. (2013). *CELF-4. Clinical evaluation of language fundamentals - fourth edition.* Bloomington, MN: NCS Pearson, Inc.

Språkens hus (n.d.) Material. Kent-Rosanoff list. www.sprakenshus.se/material.

Verhallen, M. & Schoonen, R. (1993). Lexical knowledge of monolingual and bilingual children. *Applied linguistics*, *14(4)*, 344-363.

Verhallen, M. & Schoonen, R. (1998). Lexical knowledge in L1 and L2 of third and fifth graders. *Applied Linguistics*, *19(4)*, 452-470.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, *22(2)*, 217-234.