

The role of the lexicon of school books in school achievement

Valéria Juhász

University of Szeged; juhaszvaleria@jgypk.szte.hu

Abstract

A key constituent of reading comprehension and learning ability is vocabulary. Mapping the children's vocabulary will help consciously planned developing work. In this study research on vocabulary development will be presented, and then the importance of the role of vocabulary in understanding the lexicon of texts will be introduced. The research (Nagy 2004, Nation et al. 1997) shows that knowing the 5-6000 most frequent words is sufficient to understand everyday texts, because they cover almost 95 per cent of texts. The 2008 research of the 'Vocabulary net' department shows that the schoolbook families of various publishing houses contain 40 thousand words from 3rd to 8th grades. According to the 2004 examination by Nagy (2004), children in the 4th grade can recognise 4000 words on average by reading, which rises only to 4500 by the 10th grade. It can be seen from this data that an average child can hardly keep pace with the new terminologies/lexicon of the schoolbooks. Finally, a series of books will be presented whose lexicon is in line with the children's hypothetical vocabulary, and its aim is to teach the repetitive words of the texts by using the sight word technique (but not as the method for teaching reading but as a level of reading ability, see Juhász 2019).

Keywords: *vocabulary, lexicon of school books, reading comprehension, learning abilities*

The concepts of reading

The concept of reading has changed a lot due to the research in this field (Steklács 2013). An early concept of reading grabs the aspect of knowing letters. Then, reading out meaningfully meant knowing reading. Kingston (1967) brought the communication aspect into its concept. Then, the comprehension of reading was added to the concept. Snow (2002) determines reading comprehension as the process of simultaneously extracting and constructing meaning through interaction and involvement with written language. A decade later, in PISA assessment (2015) reading was defined in a broader concept of literacy: "Reading literacy is understanding, using, reflecting on and engaging with written texts, in order to achieve one's goals, develop one's knowledge and potential, and participate in society" (PISA 2015: 49).

Predicting facts and abilities about the success of learning to read

Acquiring literacy skills needs several well working skills and cognitive processes such as speech perception and comprehension, verbal and visuospatial working memory, working intermodal relationships, executive functions, part-whole relationships, visual and verbal seriality, visual and verbal long-term memory, concentration, spatial orientation, rhythm perception and shape constancy. In addition, knowing the names of the letters and letter knowledge in kindergarten predict later reading success. Other factors such as the size of the vocabulary, language awareness and the automatic level of the recognition of grapheme-phoneme correspondences influence reading acquisition (Blomert & Csépe 2012).

A competent reader

After learning to decode reading, one must develop oneself to be a competent reader to make oneself a competent learner. But who can be referred to as a competent reader? Based on Steklács's (2013) research we can summarize it as the followings: a competent reader is someone

- who has a positive attitude to reading, who loves reading and getting information or experience from books;
- who uses what he/she has already known for what he/she is reading;
- who makes conclusions from what has been read;
- who has effective reading strategies to enhance his/her own comprehension;
- who admits what he/she understands and what they don't;
- who creates mental representations about what has been understood during and after reading;

- who asks questions in connection with the read ideas;
- who makes comparisons between the meaning of what is being read and the critical meaning of it;
- who has a rich vocabulary and good speaking skills;
- who makes connections between previous, relevant knowledge and what is being read;
- who has an effective word recognition;
- who summarizes what has been understood;
- who reads fluently enough to be able to concentrate on meaning;
- who chooses a thinking process or procedure.

From the list above it can be easily understood why all of them are crucial in meaningful reading. However, in this paper the focus is kept only on vocabulary and word recognition in schoolbooks.

Assembling the lexicon of a language or assessing the vocabulary of a person is desirable from different points of views, but as for the results they might only be approximate. On the one hand, it will be determined by the continuous changing of the number of the words, and on the other hand, by the possibilities offered by the methods and the equipment (Gósy 1999). For the teachers – from the point of view of their efficiency – it is essential to map the size of their students’ vocabulary, because it is a very important factor in predicting reading comprehension (Baker et al. 1998, Stahl & Fairbanks 1986), and in learning ability and academic performance, too. Vocabulary in itself does not guarantee understanding a text; however, its lack guarantees failure in understanding (Biemiller 2005: 223). If we know the words that are typical with a normally developing child at a certain age, we can carry out a scientifically planned and so a more effective development with those children who have a poorer vocabulary (Neuberger 2017). Knowing only the size of the children’s vocabulary is not enough to start their development, because if we are to support their academic performance, we need to monitor their vocabulary on a regular basis; that is what we need to compare with the words used in the schoolbooks, and the difference between the two will steer our developing work.

Defining the estimated size of the vocabulary

With every vocabulary assessment it has to be decided what is considered one word. All the possible forms of a word are included in the corpus (or they are considered a word family), and perhaps some inflected forms, too. Do the compound words whose meaning can be transparently deducted from the meaning of the constituting words belong to the corpus, or only those ones whose joint meaning is more than the summarised meaning of the individual words of the compound word? Making the distinction here is a problem in itself. Do function words (the, a, an, and, that, etc.) belong to the lexicon, or content words have to be counted only? What kind of differences does it cause to make a distinction here? How to decide on the phrasal verbs or the proper nouns? The research reports do not often contain these corpus arrangement details.

The vocabulary estimations of individuals published in the international and national literature are usually based on data from written materials (sometimes spoken language corpuses are included) (Crystal et al. 1998: 115). Such are e.g. writers’ dictionaries: Petőfi, a famous Hungarian poet’s dictionary contains 22,719 words, his comrade’s, János Arany’s (who translated e.g. Shakespeare into Hungarian) is estimated to have five times as many words in his vocabulary, and the Mikes dictionary based on the letters written in exile in the Ottoman Empire contains a million and a half words (Balázs 2017). Assembling dictionaries by the frequency of words is based on texts from different genres, vocabulary estimations are done according to them (Crystal et al. 1998).

Vocabulary development of individuals

In the table below there is some data connected to the growth of vocabulary and its estimated size.

Table 1: The size of the estimated vocabulary by age.

author	age	estimated size of vocabulary
Crystal (1998), Gósy (1999)	3	1,000–1,200 The least 150, the most 2,500 (Gósy 1999: 189)
Crystal (1998)	4	1,500

Crystal (1998)	5	2,000
Crystal (1998) Büky (1984) Gósy-Kovács (2001)	6	2,500 2-3,000 active Also contains data about an estimation of 8-24,000
Crystal (1998)	7	3,000
Meixner (1971)	8	260-1,468 active vocabulary
Crystal (1998)	9	4,000
Crystal (1998) Pléh (2006)	11 above 10	5,000 30-40,000 the widely read
Nádasdy (2007)	adults	3-5,000 (an average blue collar worker) 8-10,000 (smaller vocabulary together with the passive one) 50-60,000 (whole vocabulary of a literate man)
Lukács (2014)	adults	50-100,000
Gleason-Ratner (1998)	adults	75-150,000 (cited by: Gósy-Kovács 2001)

The table above shows clearly that between the estimated active and passive vocabulary the difference might be a few tens of thousands of words. The daily vocabulary building is between 3-15 words (Gósy-Kovács 2001), between the ages 1-15 the daily vocabulary building is 2-3 words (Pléh 2006), and Macher (2009) writes about 9 words a day until the age of 6.

The number of words recognised by reading by the age

Nagy (2004) assessed how many words the lower graders recognise by reading. It was compared with the results of the 10th graders.

Table 2: The number of words recognised by reading (Nagy 2004).

Time	Number of words recognised by reading
At the beginning of grade 2	3,600
At the beginning of grade 4	more than 4,000
Grade 10	4,200-4,500

These 4,000 words refer to the fact that at the beginning of the 4th grade taking the national average the students have a vocabulary of 95 per cent of the everyday text words. After the 4th grade until the beginning of the 10th grade the growth of the critical vocabulary is only 350 words according to Nagy’s (2004) assessment.

The number of the words recognised by reading is mainly determined by the number and the degree of use of the words in the student’s mental lexicon. If we compare the data by age in table 1 with the number of words recognised by reading on average, we can hypothesize that the ‘truth’ is somewhere between the data of Crystal (1998) and of Pléh (2006) (with a rather large deviation, individual differences, plus adding the difference between the active and the passive vocabulary) concerning the average vocabulary of the 10-11 year olds.

The relationship of vocabulary and comprehension

According to Crystal (1998: 115) if we examine an arbitrary text, the 100 most frequent words make up 60 per cent of the text, the 1000 most frequent words make up 85 per cent of the text, and the 4000 most frequent words make up 97.5 per cent of the text. Nation et al. (1997) refer to the research by Francis and Kucera (1982) who based on the frequency of the words in the Cobuild dictionary observed that the 1000 most frequent words make up 72 per cent of texts, the 2000 most frequent words make up 78 per cent, the 3000 most frequent words make up 84 per cent, the 4000 most frequent words make up 87 per cent, the 6000 most frequent words make up 90 per cent, and approx. the 16000 most frequent words make up 96 per cent of the everyday spoken texts (Schonell 1956). According to Laufer’s (1989) research one has to know 95 per cent of the words of the text to understand it well. According to

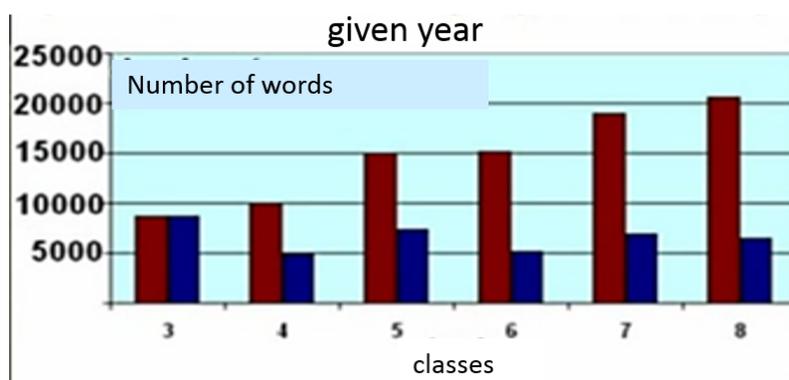
Hirsch's (1992) examination knowing the 3000 most frequent word families (roughly 5000 words) covers 95 per cent of the texts, which is sufficient to understand the text by reading (Laufer 1989, Nagy 2004).

Varga Kornél's (NYI) report on the results of the 'Vocabulary net' Department of the Hungarian Pedagogical Association changes the viewpoint of the lexicon of schoolbooks. The entire vocabulary of 99 primary schoolbooks in the Hungarian language was examined. The books were on the 2008 available schoolbook list, selected from three publishers in the main school subjects from grades 3-8. In these books 65,000 (!) basic words were identified (only verbs, nouns – except for proper nouns – and adjectives totaling 60,000 words), i.e. uninflected words. Doing the examination of the whole book series by publishers 40,000 words on average were identified.

In the graph below the number of new words is shown the children are exposed to in a school year based on the schoolbooks (blue column).

Figure 1: The number of words in a given year in the schoolbooks (Varga NYI).

Students meet 8-20000 words in one year in their school books. Blue columns mean new words in a given year



The graph shows that 5th graders come across 15,000 words in their books, out of which approx. 8,000 words are new. The students meet several thousands of new words every year while they do not come across many thousands of words they learnt in the previous years.

If we compare Varga's data with the critical vocabulary described by Nagy (2004), it can be seen that the 5,000-word critical vocabulary is rather little so that the children can understand the schoolbook texts. If we take the 40,000-word vocabulary as average, the 5,000-word critical vocabulary is only 12.5 per cent. Though little is known about how much the children's vocabulary overlaps the vocabulary of the schoolbooks, nor do we know how many of the students have an insufficient vocabulary to understand the schoolbook texts. According to the data shown in Table 1, the children's vocabulary is presumably way too below the threshold to understand the texts used in the schools (except for Pléh 2006).

Pilot research

One part of a teacher further training programme was focused on making the teachers sensitive to the schoolbook lexicon. Twenty-seven teachers took part in the training, representing different subjects; inter alia history, maths, music, Hungarian, P.E., etc. The teachers received an only two-page lesson from a 5th grader history book illustrated with pictures. Their task was to underline the words they think should be explained in the history lesson to make the children understand the text. The children were not classified into groups of gifted learners, regular learners or learners with special needs. The least number of words underlined by a teacher was nine; the most was twenty-four. These results surprised the teachers themselves, they started commenting on them, began explaining each other (without request) why they thought they were right to underline those words. Nevertheless, some consequences can be drawn from this pilot research:

- One is that in a 45-minute lesson nine unknown words are too many as children have about 6 lessons a day in the 5th grade. If they hear only half of this amount of unknown words, it means 24-30 words on a daily basis, and many of these words might not repeat at all. A week would mean 5 times more, i.e. 120-150 new words.
- The other one, the 24 unknown words mean an extreme, and that there are colleagues that are more sensitive to the children's diversity. This number of words can only be processed by children with very good abilities, and according to teachers' experience, only with the teacher's assistance.

Vocabulary building

The metaresearch by Hattie (2008) observes that out of the reading development programmes vocabulary building is the one which defines the most strongly the development of reading comprehension, before repeated reading programmes, development of visual perception, direct reading comprehension development, and letter-phoneme automation development.

It is indispensable to repeat the new words many times in different contexts to make it become part of the mental lexicon. This research makes it clear, too, that the lexicon of our schoolbooks is at least questionable.

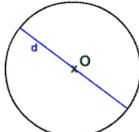
The content of the mental lexicon influences the word form recognition system, i.e. a sensible word can be recognised faster than a series of letters/words that does not make sense. It is easily understandable why it has an extreme importance how big the size of children's vocabulary is when they get down to learning to read.

Vocabulary building starts with defining what words and expressions are worth teaching the children from a given text. Selecting the lexicon cannot be done by chance. Vocabulary building must be a decision-making process which takes the following into consideration:

- How much the word in question is a representative member of the word family the student has to learn;
- How much the word can be transported, i.e. how much the student will have to use it in speaking and writing.

Before selecting the word teaching method, it has to be tracked down how often the word appears in the text, i.e. if its meaning can be deduced from the context, and if the structure of the word helps to reveal the meaning (Fisher 2016: 51). As long as the meaning cannot be deduced from the text, another word teaching method has to be chosen, e.g. mnemonics, word cards, synonyms, antonyms, descriptions, dramatization, i.e. gesturing the word or the Frayer method.

Figure 2: An example for the word card from Frayer method (Fisher 2016).

word, abstract noun: diameter	writing a definition based on experience/previous learning: A straight line passing from side to side through the centre of a body or figure, especially a circle or sphere.
picture: 	Antonym or a non-example: A straight line that touches a curve or curved surface at a point, but if extended does not cross it at that point.

The metacognitive strategy (according to the effect size of Cohen's d , $d=0.53$ Hattie 2008) is very effective in the long run as this strategy provides a technique for learning the meaning of a word. The teacher demonstrates how to think about the meaning of the word, how to try to understand the word from previous experience, from the structure of the word, or from the context.

Word categorisation works, too, when the words are put into groups based on their orthography/spelling: e.g. hardship, friendship may be in the same category as their stems end in $-d$, and the suffix is the same. But words can be grouped by other categories, e.g. teaching new words in word families also makes sense. When teaching a word, the characteristics of the word must be

examined and the possibility if the word will turn up later, perhaps in another form. Only those words should be taught directly or from a dictionary which cannot be taught by induction.

The relationship between vocabulary and reading motivation

One of the aims of teaching reading is to expose the children to the biggest number of words possible through reading. Vocabulary gained from everyday experience, i.e. not from reading experience, is relatively weak (compared to the vocabulary available from reading), and it does not provide a properly flexible framework for accepting new knowledge, adaptability to strange situations, and proper base in quality and quantity to develop problem-solving thinking.

To make this knowledge accessible, and to develop the need for a lifelong, effective and purposeful studying, it is indispensable that the children have reading motivation and that it be maintained in the long run.

A cardinal question in creating reading motivation is to give the children texts in which most words are known for them. Without this their reading will become erratic, comprehension will be slow, hard and tiring. As a result, the further reading willingness will be demotivated. In other words, it is not only necessary to give the children interesting books or texts that this way or another can be related to their interests to create reading motivation, but it is necessary to adjust the texts not to their age but to their reading age, in which the frequency ratio of the words meet the above mentioned criteria. Reading age refers to a child's reading ability after a given number of years learning to read; i.e. a child with typical reading development after three years of learning to read will be able to read on the third graders' level. An atypical child might be on the same level but in the 5th grade. In this situation this child should not be made to read textbooks for the 5th graders, but the ones that match the 3rd graders. This is why it would be very important to offer schoolbooks, compulsory reads to children on a wide scale where they could be adjusted to the reading age.

Guthrie's and Wigfield's (2000) research says a lot about the size of the vocabulary. They observed that the children who had more choices what to read, read more and more willingly, so it was obvious that their vocabulary was larger, which meant their academic performance was better, too (Anderson et al. 1988, Fisher et al. 2016). These kinds of research results motivate teachers to take books into the classrooms that are adjusted to the age and the reading age, and so make reading available for even more students. Interestingly enough Rowe (1985) (based on his own research) claims that reading comprehension developing programmes have a bigger direct effect on vocabulary development ($d=1.77$) than on the reading comprehension results ($d=0.70$), which later can of course be reconverted.

Books which help to develop sight word reading

To reach the sight word recognition reading technique, children should read a lot of books according to their reading age. The reading age means the children are offered books to read which give them sense of achievement but there are also some challenges in them. There are some series of short books which offer stories to read on various reading levels beginning with picture books and the 2nd level offers some short sentences with big pictures. The words in this series of books are repeated many times to reach sight word level soon, and at the beginning the format of the words does not change much. In Hungary one of the series I could offer for example is Aranyfa series. (Examples can be found for this: <https://en.calameo.com/read/003374024b2fbca20ed92>). These kinds of series can be found in other languages as well, e.g. <https://czytamsobie.pl/>.

Closing remarks

This paper sheds light on the fact that the 5-thousand-word vocabulary which is sufficient to understand average texts is rather insufficient for the children to understand the words in the schoolbooks. From grades 3 to 8 the 40,000 words in the schoolbooks and the too many unknown words in the teaching units make it practically impossible for an average child to have a sense of achievement in school. However, because it is our aim to educate and make our children be used to lifelong learning, creating a positive attitude towards learning is indispensable. Considering the given conditions, for the majority of students it can hardly be implemented. In order to maintain learning motivation in the long

run, besides many other factors the vocabulary of the schoolbooks has to be rethought and rediscussed by the policy makers and textbook authors.

References

- Anderson, R., Wilson, P. & Fielding, L. (1998). Growth in reading and how children spend their time outside school. *Reading Research Quarterly*, 23, 285-303.
- Baker, S., Simmons, D. & Kame'enui, E. (1998). *Vocabulary acquisition: Synthesis of the research*. Washington DC: U.S. Department of Education, Office of Educational Research and Improvement, Educational Resources.
- Balázs, G. (2017). *AranyJános's vocabulary, his poetic lexicon*. <http://e-nyelvmagazin.hu/2017/02/28/arany-janos-szokincse-koltoi-szotara/> (Accessed on 8 May 2018).
- Biemiller, A. (2005). Size and Sequence in Vocabulary Development: Implications for Choosing Words for Primary Grade Vocabulary Instruction. In E. Hiebert & M. Kamil (Eds.), *Teaching and Learning Vocabulary: Bringing Research to Practice* (pp. 223-242). NJ: Mahwah, Erlbaum.
- Blomert, L. & Csépe, V. (2012). Psychological foundations of reading acquisition and assessment. In B. Csapó & V. Csépe (Eds.), *Framework for Diagnostic Assessment of Reading* (pp. 17-85). Budapest: Nemzeti Tankönyv kiadó.
- Büky, B. (1984). The state of development and further developing the first language by age. In B. Büky, A. Egyed & C. Pléh (Eds.), *Linguistic abilities – notions – comprehension: The practical possibilities of psycholinguistics* (pp. 68-83). Budapest: Tankönyv kiadó.
- Crystal, D. (1998). *The Cambridge Encyclopedia of Language*. Budapest: Osiris.
- Fisher, D., Frey, N. & Hattie, J. (2016). *Visible learning for literacy. Implementing the practices that work best to accelerate student learning. Grades K-12*. California: Corwin Literacy.
- Francis, N. & Kučera, H. (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin.
- Gleason, B. & Bernstein, R. (Eds.) (1998). *Psycholinguistics*. Orlando: Harcourt Brace College Publishers.
- Gósy, M. (1999). *Psycholinguistics*. Budapest: Corvina Kiadó.
- Gósy, M. & Kovács, M. (2001). The mental lexicon in the view of word associations. *Magyar Nyelvőr*, 3, 330-354. <http://www.c3.hu/~nyelvor/period/1253/125306.htm>. (Accessed on 1 August 2018).
- Guthrie, J. & Wigfield, A. (2000). Engagement and motivation in reading. In M. Kamil, P. Mosenthal, P. Pearson & R. Barr (Eds.), *Handbook of reading research III* (pp. 403-424). NJ: Erlbaum. Mahwah.
- Gyarmathy, É. (2009). Cognitive Profile Test. *Iskolakultúra*, 3/4, 60-73.
- Hattie, J. A. C. (2008). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.
- Hirsh, D. & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8 (2), 689-696.
- Juhász, V. (2019). Application of the "Sight word" reading teaching method in the reading development of Hungarian pupils. In E. Berghauer-Olasz, I. Gávriiljuk, E. Hutterer & K. Pally (Eds.), *Challenges of the 21st century in the field of secondary and higher education of eastern and central Europe in light of educational reforms* (pp. 179-189). International Academic Conference Berehove, March 28–29, 2019. Rákóczi Ferenc Kárpátaljai Magyar Főiskola.
- Kingstone, A. (1967). Some thoughts on reading comprehension. In L. Hafner (Ed.), *Improving reading comprehension in secondary schools* (pp. 72-75). Macmillan, New York.
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special Language: From Humans Thinking to Thinking Machines* (pp. 316-323), Clevedon: Multilingual Matters.
- Lukács, Á. (2014). Learning words. In C. Pléh & A. Lukács (Eds.), *Psycholinguistics 1*. Budapest: Akadémiai Kiadó.
- Macher, M. (2016). Word activation and word learning processes with children that have learning difficulties. *Gyógypedagógiai Szemle*, 4, 257–269.
- Meixner, I. (1971). The vocabulary of a three year old infant. In J. Molnár & Wachal (Eds.), *The Hungarian presentations of speech symposium* (pp. 55-59). Szeged, Budapest: Nyelvtudományi Társaság.
- Nádasdy, Á. (2007). An average English worker. *Magyar Narancs*, 8 (2). http://seas3.elte.hu/delg/publications/modern_talking/107.html. (Accessed on 12 August 2018).
- Nagy, J. (2004). Criteria oriented diagnostic assessment of the development of word reading ability. *Magyar Pedagógia*, 104 (2), 123-142.
- Nation, P. & Waring, R. (1997). *Vocabulary size, text coverage and word list*. <http://www.robwaring.org/papers/CUP/cup.html>. (Accessed on 22 June 2018).
- Neuberger, T. (2017). Vocabulary building in speech development. In J. Bóna (ed.), *New methods in child language research* (pp. 121-140). Budapest: Eötvös Kiadó.

- PISA 2015 (2015). *Reading framework*. https://read.oecd-ilibrary.org/education/pisa-2015-assessment-and-analytical-framework/pisa-2015-reading-framework_9789264255425-4-en#page3. (Accessed on 23 October 2018).
- Pléh, Cs. (2006). The child language. In F. Kiefer (Ed.), *Magyar nyelv* (pp. 725-752). Budapest: Akadémiai Kiadó.
- Rowe, D. W. (1985). *The big picture: A quantitative meta-analysis of reading comprehension research*. Bloomington: Indiana University, Language Education Department.
- Schonell, F. (1956). *A study of the oral vocabulary of adults: an investigation into the spoken vocabulary of the Australian worker*. Brisbane: University of Queensland Press.
- Snow, C. E. (2002). *Reading for understanding. Toward an R&D program in reading comprehension*. Santa Monica: Rand.
- Stahl, S. A. & Fairbanks, M. M. (1986). The Effects of Vocabulary Instruction: A Model-Based Meta-Analysis. *Review of Educational Research*, 56 (1), 72-110.
- Steklács, J. (2013). *Teaching and learning reading strategies, and the belief about reading*. Budapest: Nemzedékek Tudása Kiadó.
- Varga, K. (NYI). *Vocabulary in school textbooks. The video report of 'Vocabulary net' Department about the lexicon of 99 primary school textbooks*. http://www.szokincshalo.hu/video_osszefoglalo.php. (Accessed on 2 December 2017).