

**Semantic versus Thematic Clustering of New Lexical
Items: A Textbook Analysis**

Jaakko Kujala

Bachelor's Thesis of English
Autumn term 2015
Faculty of Humanities
University of Jyväskylä

ABSTRACT

Kujala, Jaakko. 2015. Semantic versus thematic clustering of new lexical items: A school book analysis. Bachelor's Thesis of English. University of Jyväskylä. Faculty of Humanities.

Vocabulary acquisition is an essential part of learning a second language, yet textbook designers and language educators often have difficulties with creating such vocabulary acquisition exercises and materials that would best enhance the development of this crucial skill. There is indeed a growing body of evidence suggesting that semantic clustering of new lexical items impedes L2 vocabulary acquisition, but still semantic clustering is the prevalent method used in contemporary EFL textbooks. This study focused on one EFL book series in particular, which can justifiably be seen as a representative of all contemporary beginner-level EFL textbooks in Finnish elementary schools. The contents of the book were analyzed in the light of current research on vocabulary acquisition, and an alternative organizing method of new lexical items, thematic clustering, was reviewed.

Tags: Semantic clustering, thematic clustering, vocabulary teaching, English as a foreign language, EFL classroom, textbook design, materials.

Table of contents

1	INTRODUCTION.....	4
2	CLUSTERING OF VOCABULARY.....	5
2.1	Semantic clustering.....	5
2.1.1	Schema connections.....	5
2.1.2	Semantic field theory.....	6
2.1.3	The convenience hypothesis.....	6
2.2	Thematic clustering.....	7
2.2.1	Interference theory.....	7
2.2.2	The distinctiveness hypothesis.....	8
3	RESEARCH DESIGN.....	9
3.1	Research questions.....	9
3.2	Research progress.....	9
3.2.1	Selection of material.....	9
3.2.2	Selection of exercises.....	10
3.2.3	Categorizing clusters.....	10
3.2.4	Exclusion of e-material and Teacher's Material.....	11
3.3	Data analysis.....	11
4	DISCUSSION.....	13
4.1	Convenience over results?.....	13
4.2	Quantification of acquisition.....	14
4.3	Correlations to learner level.....	15
4.4	Absence of long-period studies.....	15
4.5	Thematic textbook design.....	15
5	CONCLUSION.....	16
6	SOURCES.....	17

1 INTRODUCTION

There is a growing consensus among English Language Teaching (ELT) researchers that lexical knowledge is an essential component of speakers' communicative competence (discussed in Basanta 2010). Therefore, language educators and education material designers need to assess and reassess the methods of vocabulary instruction in order to facilitate the best possible vocabulary learning results. Traditionally, EFL school books as well as language teachers have presented new lexical content in semantically preorganized patterns called semantic clusters. This tradition is rooted in some psycho-linguistic findings about L2 learners' mental lexicon, which seems to be semantically organized (Lehrer 1974). There is, however, a growing body of empirical evidence (Erten and Tekin 2008, Finkbeiner and Nicol 2003) suggesting that semantic clustering actually hinders vocabulary acquisition and retention among novice L2 learners, which confirms Higa's early theory of interference. Higa has shown that memory traces compete with each other, and his 'Interference theory' suggests that if new lexical items are to be presented to learners, they should not be presented in word clusters that share a common head word or superordinate concept (Higa 1963).

The purpose of this study is to investigate the most commonly used EFL textbook for third graders in Finnish elementary schools (according to publishers Otava and SanomaPro, and the Federation of Foreign Language Teachers in Finland (SUKOL)) called *Yippee! 3*, as a representative of Finnish EFL school books for beginner learners, and to see whether it is designed according to the semantic clustering tradition when it comes to presenting new words. This may be of great help for EFL educators when they are in the process of choosing vocabulary instruction materials for their classes. In addition, material designers might find this helpful when reassessing their work.

2 CLUSTERING OF VOCABULARY

2.1 Semantic clustering

Semantic clustering means that new lexical items are presented for language learners in word sets that share a common field of meaning (Tinkham 1997). Semantic clusters could be, for example, *colors*, *household appliances* or *family relatives*. The words in semantic clusters usually are not only of the same semantic field but also of the same word class like adjectives (colors) or nouns (household appliances and relatives). Researchers in the field of semantic clustering acknowledge that although there is some empirical evidence in favor of this method from L1 language teaching, the evidence L2 is almost non-existent (Wilcox & Medina 2013). Next, however, we will look into some of the theory contributing to semantic clustering.

2.1.1 Schema connections

Presenting new words in clusters, i.e. sets of words, in EFL classrooms and education material has been a given method, since instruction is usually divided into rather brief lectures with limited time allotment, which makes individual word presentation unmeaningful. The idea of presenting new words in semantically-related sets, i.e. semantic clusters has its origins in educational psychology. Probably the earliest and most influential proponent of semantic clustering is psycholinguist Ausubel. He stated (Ausubel 1968) that superordinate concepts, under which words are organized, should be presented in advance in order to activate the existing schema in the mental lexicon, which prepares the organization of new lexical items into these pre-activated slots.

2.1.2 Semantic field theory

The tradition to present new words in EFL learning settings particularly in semantic clusters also rises from Lehrer's semantic field theory (Lehrer 1974), which is likewise based on psycholinguistics. Semantic field theory states that vocabulary is cognitively organized in interrelated connections between lexical items in the human mind. The mind classifies words by their meaning and forms connections between them, and these networks of connections are called semantic fields.

Channell (1981) further elaborated Lehrer's theory and stated that semantically closely related words are actually located physically near each other in the mental lexicon. These findings have led educators and material designers to conclude that vocabulary should be presented in semantic clusters to foster the natural progress in the semantically organized mental lexicon (discussed in Wilcox and Medina 2013).

2.1.3 The convenience hypothesis

Regarding the almost non-existent empirical evidence in favor of semantic clustering Tinkham (1997) argues that the prevalent usage of semantically organized word sets results rather from convenience than from any well-grounded theory basis. Firstly, organizing vocabulary in semantic clusters is convenient for the language educators because they will have to cover some pre-chosen areas of the national curriculum with their classes, and by presenting words in semantic clusters it is easy to supervise whether the areas concerning vocabulary instruction are covered in due time. Consequently, material designers strive to produce material which is organized in semantic clusters, since that kind of material serves the needs of many educators in the market.

Another aspect in favor of the convenience hypothesis is the design of many L2 exercises. A very common concept in L2 instruction is the filling-the-gaps exercise. This kind of single word recognition exercise types support the presentation of new vocabulary in semantically and syntactically related sets. When, for example, the clause to be filled in goes like this: " I bought a _____

jacket” (followed by the L1 equivalent of the wanted word), it is very convenient to present the new words in semantic clusters, which in this case would be, for example, colors or clothing materials. Convenience, indeed, seems to be the only rational reason to present new vocabulary for L2 learners in semantic clusters. Next we will look into another option of presenting words in clusters.

2.2 Thematic clustering

The concept of thematic clustering means that new words are presented for learners in thematic sets instead of semantic sets. Whereas a semantic cluster would, for example, contain words like *green*, *blue*, *red* and *yellow*, a thematic cluster would contain words like *green*, *frog*, *hop* and *pond*. The words in the semantic cluster carry the similar semantic meaning (colors) and the same syntactic form (nouns), whereas the words in the thematic cluster vary with their semantic content and are of different word classes (adjectives, nouns and verbs) but are still related thematically to the concept of the frog and its environment and behavior.

2.2.1 Interference theory

Higa’s theory of interference (Higa 1963) is probably the earliest opponent of semantic clustering. Interference theory states that if semantically similar words are presented simultaneously, it will interfere the learner’s vocabulary acquisition and lexical retention since the semantic content of the words are quite indistinguishable (Higa 1963, Waring 1997). Words in semantic clusters not only share similar semantic content but are of the same word class, which adds to the interference effect. Tinkham (in Tinkham 1997) further stated that the interference of semantic clusters is especially prevalent among novice L2 learners, since their lexical knowledge is yet poorly established.

There is a growing body of empirical evidence supporting interference theory (Finkbeiner and Nicol 2003; Schneider et al. 2002; Papathanasiou 2009; Erten and Tekin 2008; Wilcox and Medina 2013). These studies clearly conclude

that presenting new vocabulary in semantically related sets decreases vocabulary acquisition and has negative long-term effects on vocabulary retention when compared with presenting words in semantically unrelated sets.

2.2.2 The distinctiveness hypothesis

The distinctiveness hypothesis, following Higa's interference theory, suggests that since the similarity of newly presented vocabulary hinders learning of new words, lexical items presented should be as distinct as possible – the less similarity between words, the easier is the acquisition and retention of the words (Hunt and Elliot 1980, Hunt and Mitchell 1982). According to this hypothesis words should be presented from a variety of word classes (nouns, verbs, adjectives etc.) and different semantic contents in order to facilitate vocabulary acquisition and retention.

Distinctiveness on the level of orthography and phonology also increase word recognition and thus facilitate acquisition and retention rate (Wilcox and Medina 2013). The distinctiveness hypothesis does not however suggest that the words should be unrelated or arbitrarily selected. This hypothesis actually supports selecting words that share thematic relationships. Tinkham (1997) compared learners' acquisition and retention rate of semantic clusters, thematic clusters and arbitrarily selected words, and the results of his study clearly showed that the most beneficial for learning were thematic clusters. Semantic clusters clearly impeded word acquisition and retention ending in poorer results than arbitrarily chosen word sets.

3 RESEARCH DESIGN

3.1 Research questions

In this study the following two research questions will be addressed:

- 1a) To what extent, if any, has the *semantic clustering* of vocabulary been used when presenting new words in the Finnish EFL textbook, *Yippee! 3*?
- 1b) To what extent, if any, has the *thematic clustering* of vocabulary been used when presenting new words in the Finnish EFL textbook, *Yippee! 3*?
- 2) What are the underlying reasons behind the usage of these clustering methods?

3.2 Research progress

3.2.1 Selection of material

I chose the book series *Yippee!* since it is the most commonly used EFL book series in Finnish primary schools in the academic year 2015-2016, and thus can be seen as a representative of Finnish school books. This information was gained by consulting the two leading publishers in the market of school books, Otava and SanomaPro, as well as the Federation of Foreign Language Teachers in Finland (SUKOL). Although the exact sales records are classified documents, owned by the publishers, there was a consensus among the aforementioned sources that *Yippee!* is the most popular series of this year and is thus qualified to represent the mainstream of EFL instruction materials in Finland.

I chose *Yippee! 3* since it is the first book of the series and is designed for third graders. This was an important criterion because most Finnish primary school pupils start learning English in the third grade (according to SUKOL) and therefore can represent the category of novice EFL learners, with whom the difference between semantic and thematic clustering is the most significant. The Finnish third graders are at the age of 8 to 10, which is a critical phase for L2 acquisition, which gives a lot prestige for vocabulary teaching materials for that school grade.

3.2.2 Selection of exercises

There were two criteria for selecting exercises to analyze in this study:

- 1) **The main function of the exercise was be vocabulary acquisition.** This was in contrast with exercises focusing on other skills, such as: pronunciation, grammar, reading comprehension and listening comprehension.
- 2) **There was a cluster of words presented.** Here a cluster meant a set of words containing at least three words. The cluster could be in any form in the exercise: in written or spoken words or pictures.

The total number of exercises in *Yippee! 3* that met the afore-mentioned requirements was 175 exercises. These were selected both from the textbook and the exercise book, or *Yippee! 3 Reader* and *Yippee! 3 Writer*.

3.2.3 Categorizing clusters

Each one of the exercises was placed under into one of the following categories according to the type of clustering of that particular exercise:

- 1) **Semantic clusters.** I.e. sets of words that share a closely related semantic meaning, such as *colors* or *vehicles*.
- 2) **Thematic clusters.** i. e. sets of words that share a closely related thematic meaning, such as *frog, hop, green* and *pond*.

- 3) **Non-related clusters.** This category contains those sets of words that share neither semantic nor thematic content. Syntactic relations were not taken into account with this category.

After the categorization of word clusters, each exercise with either semantic, thematic or non-related cluster was categorized according to the type of exercise; e.g. translations, fill-in-the-gaps, repetition and picture recognition.

3.2.4 Exclusion of e-material and Teacher's Material

It is worthwhile mentioning that *Yippee! 3* course has its own e-material with an abundance of additional exercises for learners. I, however, excluded this e-material from my study since the exercises provided in this web-based learning environment were to great extent rather similar and on some occasions even identical to the workbook exercises. Including e-material to this study would not have carried any added value to the study, and thus it was reasonable to exclude it. In addition, *Yippee! 3* comes with a teacher's material consisting of a proposed example syllabus, additional exercises, help tools and means of differentiating teaching. These additional features would neither have given any added value to the study nor been accessible to learners, and thus they were excluded from the data.

3.3 Data analysis

Table A below shows the distribution of clusters in both *Yippee! 3* workbook and textbook into three categories: semantic, thematic and un-related clusters. The category of semantic clusters was by far the most frequent, and covered nearly 80 percent of all clusters, and consisted of 17 different exercise types. Un-related clusters were the second most frequent category, covering slightly over 20 percent of all clusters and consisting of seven different exercise types. Thematic clusters were totally absent in this data and were not represented in any of the 175 exercises of this EFL material. These clusters were divided into the

three categories according to the definitions of Tinkham (1997), which were explained earlier in the discussion of theories (see 2.1 and 2.2).

SEMANTIC CLUSTERS	THEMATIC CLUSTERS	UNRELATED CLUSTERS
28x Find words		19x Repetition
23x Glossary boxes		9x Crosswords
19x Crosswords		3x Find the words
12x Repetition		2x Fill in the gaps
11x Connect word and picture		1x Connect word and picture
8x Listening comprehension		1x Circle the words you know
7x Write down all the words you know		1x Labyrinth
6x Draw the words		
6x Picture recognition		
5x Fill-in-the-gaps		
4x Colouring		
4x Labyrinth puzzle		
2x Translation		
1x Discussion		
1x Connect the dots		
1x Memory game		
1x Bingo		
Total: 139	Total: 0	Total: 36

(Table A: The categorization of *Yippee!* 3 exercises into three categories of clusters)

Interestingly, more than 50% of the unrelated clusters were presented in some identical repetition exercises. This exercise type, at least in the context of *Yippee! 3*, means that learners repeat out loud a list of words, usually related to the vocabulary of the current textbook chapter, either after the example of the teacher or of the recorded tape. These words were un-related both semantically and thematically like in the following cluster: *a heart, a computer, a surprise, a pet, fantastic* (*Yippee! 3* Writer p.56). The interesting fact here is that although semantic clusters covered over 80 percent of all clusters in this data, there were 19 repetition exercises in unrelated cluster whereas semantic clusters only had 12 repetition exercises.

4 DISCUSSION

4.1 Convenience over results?

Although thematic clustering of words has been proven by vocabulary acquisition scholars a better way to present new lexical items than semantically clustered words (see discussion in 2.2), this study showed that semantic clustering is still prevalent in Finnish EFL school books for third graders, *Yippee! 3* being the representative. Arguably the most rational reason behind this form of presenting words is Tinkham's convenience hypothesis, which states that semantic clustering is not based so much on empirical studies on vocabulary acquisition as it is based on convenience (Tinkham 1997). Semantic clusters are indeed convenient for language educators since they offer a clear schedule for the vocabulary teaching curriculum. With the help of semantic clusters EFL teachers can easily see which lexical categories are already covered and which are yet to be studied with the group of learners. Furthermore, text book designers pursue to meet the needs of language educators and national curricula, which is ultimately the reason why textbooks are constructed using semantic clustering, because of its convenience for curricula.

Language educators should naturally give more prestige to pedagogical findings than to convenience, but empirical reality suggests otherwise. Considering teachers' demanding work of managing many classes at the same time and working not only as a teacher of a subject but also to great extent as an educator, the limited and occupied resources make teachers go the easy way, which in this case is semantic clustering. Semantic clustering is something that language educators are used to, which is why it is convenient. It would need a long period of determined transition to make thematic clustering the new convenience, which ultimately should start from the National Board of Education making thematic clustering a part of the new national curriculum.

4.2 Quantification of acquisition

Another reason behind the prevalence of semantic clustering in this study could have been the concept of quantification of acquisition. Semantic clusters indeed make it easy for learners to quantify their vocabulary acquisition, and thus learners can monitor their own process of language learning. An EFL learner can, for example, look back at the previous chapters of the textbook to quantify how many semantic categories they have learned and to qualify how well these categories are indeed learned. This might greatly increase one's learning motivation. A knock-back to this hypothesis is, however, the empirically proven fact that semantic clustering impedes the learning of new lexical items. Furthermore, thematic clusters can as well be useful in quantification – one only has to use different applications to do so. Quantification of thematically learned lexical items would be more accessible if learning materials were designed in favour of thematic clustering.

4.3 Correlations to learner level

My study focused particularly on material for low-level learners. While learning is a process, always moving towards more and more applying use of language, it is rather obvious that materials for low-level learners are designed as simplistically as possible, which usually results in semantic clustering. I would argue, however, that simplicity does not have to lead to semantic clustering but it could be reached by thematic clustering as well. There is no evidence whatsoever that the simple word-meaning relations are learnt better when using semantic clustering than when using thematic clustering (Waring 1997). Thematic clustering is, according to research, superior in every level of language skills when compared to semantic clustering (Wilcox & Medina 2013).

4.4 Absence of long-period studies

The studies conducted on semantic and thematic clustering on vocabulary acquisition have so far been short-period studies, and there are no long-period studies done on thematically clustered curricula. Although learners might benefit from thematic clustering when it comes to short period testing, there is little evidence of the benefits of thematic clustering on long periods of vocabulary acquisition. Further research is needed to show whether thematic clustering is a better framework to build a curriculum upon than semantic clustering (Wilcox and Medina 2013.)

4.5 Thematic textbook design

The absence of examples in this case makes it rather difficult to picture what a thematically organized textbook would look like. The long tradition of semantic clustering has formed our conception of textbook and exercise design, and therefore it would need a long process to introduce thematically organized textbooks into regular use in EFL classrooms. Further research and considera-

tion is also needed when designing thematically organized textbooks and materials.

5 CONCLUSION

This study showed that low-level Finnish EFL textbooks are designed by the paradigm of semantic clustering instead of thematic clustering. Theories, research and contemporary findings in the field of vocabulary acquisition were reviewed and compared, and in the light of research thematic clustering has been proven a pedagogically superior method over semantic clustering. Many underlying reasons behind the textbook designers' and teachers' preference of semantic clustering were discussed, and probably the most likely reason is convenience – following the given traditions in the face of limited resources. (Tinkham 1997).

There are still however myriads of questions to be answered in this discussion between different types of clustering methods. Further research is needed to show whether thematic clustering has any long-period effects on vocabulary acquisition, for example in the time period of a whole elementary school EFL curriculum. The design of thematically organized textbooks and other learning materials need to be reconsidered, and preliminary examples need to be evaluated by vocabulary acquisition researchers. Even all these questions unanswered, it is quite clear that semantic clustering will and should be left to the history of vocabulary acquisition and new methods will eventually replace it in language learning curricula.

6 SOURCES

- Ausubel, D. 1968. *Educational psychology: A Cognitive View*. Holt, Rinehart & Winston Inc., New York.
- Basanta, C. P. 2010. A Second-Generation CALL Vocabulary-Learning Program ADELEX: In Search of a Psychopedagogic Model. *Insights into Non-native Vocabulary Teaching and Learning* (2010), pp. 175-185.
- Channell, J. 1981. Applying semantic theory to vocabulary teaching. *ELT J.*, 35 (2) (1981), pp. 115-122.
- Erten, I. H. and Tekin, M. 2008. Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System: An International Journal of Educational Technology and Applied Linguistics*, 2008, Vol.36(3), p.407-422.
- Finkbeiner, M. and Nicol, J. 2003. Semantic category effects in second language word learning. *Applied Psycholinguistics*, 24 (3) (2003), pp. 283-369.
- Higa, M. 1963. Interference effects of intralist and word relationships in verbal learning. *J. Verbal Learning and Verbal Behaviour.*, 2 (1963), pp. 170-175.
- Hunt, R.R. and Elliot, J.M. 1980. The role of nonsemantic information in memory: orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General*, 1980, Vol.109(1), pp.49-74.
- Hunt, R.R. and Mitchell, D.B. 1982. Independent effects of semantic and nonsemantic distinctiveness. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 1982, Vol.8(1), pp.81-87.
- Lehrer, A. 1974. *Semantic Fields and Lexical Structure*. North-Holland, Amsterdam.
- Papathanasiou, E. 2009. An investigation of two ways of presenting vocabulary. *ELT Journal*, 2009, Vol.63(4), p.313-322.
- Schneider, V. Healy, A. and Bourne, L. What is learned under difficult conditions is hard to forget: contextual interference effects in foreign vocabulary acquisition, retention and transfer. *Journal of Memory and Language*, 2002, Vol.46(2), pp.419-440.
- Tinkham, T. 1997. The effects of semantic and thematic clustering on the learning of a second language vocabulary. *Second Language Research*, 1 April 1997, Vol. 13(2), pp. 138-163.

Waring, R. 1997. The negative effects of learning words in semantic sets: a replication. *System*, 1997, Vol.25(2), pp.261-274.

Wilcox, A. and Medina, A. 2013. Effects of semantic and phonological clustering on L2 vocabulary acquisition among novice learners. *System*, 2013, Vol.41(4), pp.1056-1069.