



Learning vocabulary in semantic sets: Myths about myths

Zorana Vasiljevic¹, Bunkyo University, Japan

Abstract:

One important decision language teachers and writers of pedagogical materials need to make concerns the presentation and organization of the target vocabulary. While many instructors and EFL textbooks have opted for semantic clustering of new vocabulary, there is now research evidence to suggest that semantic grouping causes interference and potentially impedes the learning process. Consequently, several recent reference books on vocabulary teaching argue that new words should be grouped thematically rather than semantically. The present study was designed to test the validity of these claims. Comparisons were therefore made between the productive and receptive learning of L2 vocabulary when the target words were grouped semantically, thematically, and when the grouping was based on formal similarity. The results of immediate and delayed post-tests showed that words presented in a semantic set were recognized and recalled equally well or better than words grouped thematically. The words with similar forms were most difficult to learn, with an interference effect being particularly strong for recognition of the target vocabulary. The results do not suggest that semantic vocabulary grouping is detrimental to learning and more research is therefore needed on the impact the organization of target words has on L2 vocabulary acquisition.

Keywords: *semantic clusters, thematic clusters, semantic interference, synforms (similar lexical forms)*

Apstrakt:

Jedva važna odluka koju nastavnici stranog jezika i pisci pedagoških materijala treba da donesu tiče se predstavljanja i organizacije vokabulara koji treba da se nauči. I dok se mnogi instruktori i udžbenici engleskog kao stranog jezika odlučuju za semantičko grupisanje novog vokabulara, sada postoje dokazi u istraživanjima koji sugerišu da semantičko grupisanje dovodi do interferencije i potencijalno smeta procesu učenja. Zbog toga, u nekoliko novijih relevantnih knjiga o predavanju vokabulara tvrdi se da nove reči treba grupisati tematski a ne semantički. Ovo istraživanje sprovedeno je da bi se testirala validnost tih tvrdnji. Stoga su napravljena poređenja između produktivnog i receptivnog učenja vokabulara stranog jezika kada su reči grupisanje semantički, tematski, i kada je grupisanje bilo na osnovu sličnosti u formi. Rezultati testova sprovedenih odmah i malo nakon učenja pokazali su da su reči koje su bile predstavljene u semantičkim grupama bile podjednako dobro ili bolje prepoznavane i da su ih se učenici prisećali isto kao ili još bolje od reči koje su bile tematski grupisane. Najteže je bilo naučiti reči koje su imale sličnu formu, i efekat interferencije bio je naročito jak pri prepoznavanju targetiranog vokabulara. Rezultati ne potvrđuju da semantičko grupisanje vokabulara ima loš efekat na učenje, i stoga je potrebno još istraživanja na temu uticaja koji organizacija targetiranog vokabulara ima na usvajanje reči u stranom jeziku.

Ključne reči: *semantičko grupisanje, tematsko grupisanje, semantička interferencija, sinformi (slične leksičke forme)*

¹zornavasiljevic@yahoo.com

1. Introduction

1.1 Background

In many EFL textbooks the target vocabulary is presented in semantic sets. New words are often clustered in groups that share certain semantic and syntactic properties. Some typical examples are groupings such as *family members, colours, days of the week, occupations, fruits, or animals*. This practice is grounded in a widespread belief among coursebook writers that presenting new vocabulary in semantic clusters facilitates vocabulary building. Some evidence in favor of this approach comes from studies of human cognition. For instance, conceptual categorization enables humans to make sense of their everyday world. People strive to organize information into meaningful sets so that they can build a structured and coherent representation of the world in which they live. This also influences their use of language. Common sense relations such as *synonymy, antonymy, hyponymy* or *meronymy* are believed to reflect the human endeavor to make sense of infinitely and continuously varied life experiences (Cruse, 1986). Conceptually-related words tend to be remembered as a group and their recall improves significantly when items are organized systematically (Bousfield & Cohen, 1955; Cofer, Bruce, & Reicher, 1966). This positive effect of semantic grouping on recall memory has been observed both in studies where the order of stimuli presentation was experimentally manipulated and experiments in which the sources of stimuli organization could not be readily specified. The results of such studies provide strong evidence to show that people engage naturally in the clustering and hierarchical structuring of information (Jenkins & Russell, 1952; Tulving, 1962).

These findings have important implications for the way in which second languages are taught. If the mental lexicon consists of a series of structured, interconnected networks, then grouping words in semantic sets at the initial presentation stage may facilitate their integration in long-term memory, thus making the learning process more efficient. This argument has been proposed by several scholars. For example, Channell (1981) states that the human mind uses semantic similarity to classify words and, therefore, presenting vocabulary in semantic clusters can benefit the learning process. Seal (1991) argues that learning one lexical item can promote the learning of another. According to Marzano and Marzano (1988), semantic clusters help bring structure to instruction, establish a frame of reference for new words, and provide implicit clues about word meanings. This means that the process of word integration can begin even when no contextual clues or definitions are available. By encouraging learners to think beyond the isolated words and take advantage of prior knowledge, teachers can maximize the effectiveness of explicit instruction. Neuner (1992) contends that the learning burden is reduced when words are learned in sets. Learning a pair of antonyms such as *dead-alive* does not require any more energy or concentration than learning just one of the words. The effective use of sense relations can therefore facilitate the learning process.

Semantic grouping can also deepen learners' understanding of words that are already in their lexicon (Channell, 1981; Marzano & Marzano, 1988; Neuner, 1992). Word knowledge entails more than a knowledge of simple definitions; it comprises different sets of associations evoked by particular words.

The meaning of individual lexical items can only be fully understood in relation to other words (Cruse, 1986; Dunbar, 1992; Gairns & Redman, 1986). Componential analysis of individual words within a semantic cluster can highlight the differences between partial synonyms and help learners define their boundaries, leading to greater variety and precision of expression (Channell, 1981).

Despite the arguments above, some concerns have been raised in recent years about the possible interference effect that presenting words in semantic sets may have on vocabulary acquisition in second language learners. Reference books for language teachers (e.g., Folse, 2004; Nation, 2001; Webb & Nation, 2017) warn against semantic clustering of lexical items on the grounds that the concurrent presentation of words similar in meaning may confuse learners and increase their learning burden. Nation (2001) states that presenting formally similar words, opposites, synonyms, free associates, or co-hyponyms together makes learning more difficult. Folse (2004) argues in favour of the thematic grouping of novel lexical items. He contends, for example, that *Tuesday* and *Thursday* are often confused by ESL learners because they are both days in the middle of the week that begin with *T*. He argues that they should be taught separately, with *Tuesday* taught first as it occurs twice as often in everyday usage. Webb and Nation (2017) also contend that related items such as *days of the week* should be taught one at a time according to need or their frequency in the target language.

The arguments above are based on *interference theory*, according to which an interaction between the new material and previously learned information may have a negative effect on the speed of learning and memory performance. Interference can be *proactive* or *retroactive*, that is, older memories may inhibit the creation of new memories, or newly learned information may impede the recall of previously learned information. Interference was found to be primarily a function of similarity and the associative strength between new and old information - the more similar the material is, the more likely it is that some information will be forgotten (Crowder, 1976; McGeoch & McDonald, 1931; Underwood, 1957).

The interference effect has been widely researched in psychology and most of the evidence comes from studies conducted with native speakers. In 1963, Higa investigated interference effects for six kinds of semantic relationships: (1) synonymity, (2) antonymity, (3) free-associates (primary associates of high associative strength), (4) partial-response-identity words (words similar in terms of their free associates), (5) connotation words (words whose connotative meanings are related), and (6) coordinates. The conditions were compared by counting the number of trials the participants needed to memorize the words. Free associates, synonyms, and antonyms were found to be significantly more difficult to learn than words from the other three categories, with the greatest degree of interference occurring between near synonyms (e.g., *fast-rapid*, *tall-high*). Higa concluded that interference is a function of both, similarity in denotative meanings and associative strength. Based on these findings, Higa (1965) conducted a follow-up study that repudiated the semantic clustering of new vocabulary in second language textbooks and classroom instruction.

In 1993 and 1997 Tinkham conducted two carefully controlled studies in which he examined the effects of different patterns of word clustering on the learning of L2 vocabulary. The participants were asked to learn lists of translation pairs, consisting of an English word and an artificial “new” word, in as few trials as possible. The results showed that semantically clustered words were more difficult to learn than words grouped thematically, or vocabulary presented in semantically unrelated sets. Based on these findings, Tinkham concluded that the semantic clustering of novel vocabulary has an inhibitive effect on learning L2 vocabulary due to mutual interference between the target items.

Tinkham’s studies were replicated by Waring (1997) and Williams (2003). The results of both studies supported his findings in that lists in which the target items were semantically related took more time to learn than the lists that contained semantically unrelated items or thematically grouped words.

Concerns about the negative effect of semantic categorization were also raised by Finkbeiner and Nicol (2003). Their study examined receptive and productive knowledge of 32 pseudo words grouped in four semantic categories: *animals*, *kitchen utensils*, *furniture*, and *body parts*. The target words were presented on the computer screen together with their corresponding pictures. Retention and recall of the words was measured through translation tests. The results showed that translation times were longer and test scores were lower when the items were grouped in semantic sets.

The results of the studies reviewed above seem to consistently suggest a negative effect of semantic clustering on word retention. However, whether these findings really constitute evidence against the semantic grouping of new vocabulary in foreign language instruction remains open to question.

1.2 Issues of concern

First, all the studies reported above were conducted in strictly controlled experimental conditions. This limitation was acknowledged by some of the authors. Waring (1997), for example, admits that a tightly controlled experimental design benefits the researcher, not the learner. The studies become easier to replicate but are perhaps less applicable to classroom contexts.

Second, the target items used in the experiments were either artificial words (Tinkham, 1993, 1997; Waring, 1997; Finkbeiner & Nicol, 2003) or came from a natural language about which the participants had no prior knowledge (Williams, 2003). Some studies suggest that the interference effect may depend on the type of stimulus material used. Haberlandt (1994) states that when meaningful passages are used in place of nonsense syllables, memory traces are not affected. Furthermore, languages have different phonotactic and graphotactic rules and therefore experiments in which target words came from a language entirely new to the participants may not be indicative of their performance in a real classroom.

The use of pseudo words may have also affected participants’ motivation and their perceptions of the task. Finkbeiner and Nicol (2003) reported that 23 out of 47 participants in their study failed to reach

the predetermined accuracy of 80% and therefore their data were excluded from further analysis. The authors attributed the participants' poor performance to a lack of incentive to engage with learning an "alien language".

Both Waring (1997) and Williams (2003) reported that the participants in the experiments felt that semantic sets were more difficult to learn than other types of word clusters. However, it should be noted that both studies involved a relatively small number of participants – 20 in Waring's study and 24 in Williams's experiment. Therefore, when Williams reports that "One fourth of the participants reported that the thematic set was easiest to learn, whereas only one sixth claimed that it was difficult" (2003, p. 55), the real difference is six versus four learners. Learners' beliefs are certainly an important factor in second language vocabulary acquisition. A word is difficult if it requires a great deal of effort on the part of the learner. In other words, it is *the learner's experience* that defines what is difficult in the learning process. However, considering the small number of participants in the studies and the fact they were expected to memorize sets of nonsense words, participants' perceptions of the task can hardly be considered indicative of EFL learners' attitudes towards the semantic clustering of novel vocabulary. The only formal study that has examined learners' attitudes to presenting new vocabulary in semantic sets was Ishii and Maruyama's (2009) survey of 543 Japanese university students. Their results showed that 64% of participants believed that semantic grouping facilitated learning, while 24% were unsure. Only 12% of the respondents did not share this conviction. Although learners' beliefs can be shaped by past learning experiences (Little, Singleton, & Silvius, 1984) and do not necessarily imply an understanding of the most effective learning methods, there is no doubt they affect learners' attitudes towards classroom practices and their expectations regarding the teacher. Therefore, dismissing these beliefs as erroneous without sufficient evidence may have a negative effect on learning outcomes.

There are also some concerns regarding the criteria used for the formation of semantic sets. The majority of studies reported above examined the interference effect between coordinates (co-hyponyms). While coordinates are known to have strong associative bonds (Aitchison, 1994), semantic links are not limited to hyponymic relations. Semantic grouping based on other sense relations such as synonymy or antonymy may have produced different kinds of results. In those studies, that did examine an interference effect for different kinds of sense relations, the criteria used to categorize the target words were not always clear. For example, in Higa's (1963) experiment, *bitter* and *sour* were classified as partial-response identity words but could just as easily be placed in the category of coordinates, where *taste* would be a superordinate term. *Eagle* and *bird* were analyzed as a free-association pair; however, *flower* and *rose* were classified as connotation words. Therefore, although Higa reports the norm lists on which classifications were based, the boundaries between the categories seem somewhat arbitrary. Furthermore, in some of the studies reviewed, the target words were not only semantically related but also phonologically similar. For example, the semantic set in Williams's (2003) study included items such as *maprow* (coconut), *malagaw* (papaya), and *mamooang* (mango). While semantically related words may sometimes share phonologies, this is by

no means an intrinsic feature of semantic grouping. Given that the target words were presented aurally it is possible that phonological similarity between the target items may have been an unmeasured confounding variable. Furthermore, the target sets sometimes included words whose referents were not only semantically connected, they also shared several physical features. For example, one of the sets in Tinkham's study contained metals such as *tin*, *bronze* and *brass* which may be visually difficult to distinguish, while the set in Williams's study contained fruits like *mango* and *papaya* which are both oval shaped and of a yellow-orange colour. Some recent studies (e.g., Ishii, 2015, 2017) suggest that semantic clustering has an impeding effect only when the referents of the grouped words share physical components. This is due to difficulties in the visual mental coding of input that takes place naturally during information processing. When clusters consist of words that do not share visually similar referents, semantic grouping appears not to inhibit learning.

In some studies (e.g., Finkbeiner & Nicol, 2003) vocabulary treatment involved associations between novel words and pictures rather than L1-L2 lexical pairs. While visual input is generally believed to have a facilitative effect on information retention and recall (see Paivio, 1969), the results of several studies (e.g., Boers, Piquer, Stengers, & Eyckmans, 2009) suggest that images may have a negative effect on the retention of word forms, especially among visual learners.

Finally, there are also issues regarding the instruments used to measure word learning. Apart from Finkbeiner's and Nicol's (2003) study, where vocabulary retention was assessed through bidirectional translation tests, the studies reviewed above used trials-to-criterion tests where conditions were compared by counting the number of trials the participants needed to memorize particular sets of words. All the words in a set had to be recalled correctly for a condition to be met. However, in a real classroom, vocabulary tests are used to measure *the number of words* learners have acquired and the expectation is that some words within a cluster will be learned faster than others. Furthermore, the experiments reported above only included immediate post-tests where productive word knowledge was assessed. Even in L1, receptive vocabulary is much larger than productive vocabulary, and therefore it can be assumed that fewer trials would have been needed if word recognition tests had been employed. Other studies (e.g. Aksoy, 2014; AlShaikhi, 2011; Davies, 2012) showed that the advantages of semantic clustering are observed on delayed but not immediate post-tests. It is possible that although learning vocabulary in semantic sets is initially more difficult, formed memory traces are less susceptible to forgetting. This hypothesis would be in line with Craik and Lockhart's (1972) *Depth of processing theory* which posits that deeper levels of mental processing result in more durable memory traces.

Overall, laboratory-based experiments appear to suggest that presenting new words in groups of co-hyponyms can make learning more difficult. However, despite the meticulousness with which these studies were conducted, the extent to which their results are applicable to second language learner population is questionable. While many processes in second language acquisition are psycholinguistic in nature, the language classroom is not a psycholinguistic lab. Languages are not

learned in front of computer screens, and the semantic grouping of words in EFL materials does not normally imply the decontextualized memorization of L1-L2 translation pairs. Experimental conditions enable researchers to control variables, isolate a specific phenomenon, and replicate research more easily, but they do not reflect the complexities of a real classroom where multiple variables are always at play. Psycholinguistics can help applied linguists understand some of the mechanisms involved in language acquisition and usage; however, decisions about interventions to improve or possibly change these mechanisms must be based on data from real learners and real classrooms. In terms of the effects that semantic clustering may have on L2 vocabulary learning, this type of data is both insufficient and inconclusive. Few studies have examined how the organization of novel vocabulary affects learning and those that have been conducted yielded mixed results. These studies will be reviewed in the next section.

1.3 Classroom-based research

Data from classroom-based research are mixed and inconclusive. The arguments in favour of semantic grouping include: (1) better retention of the target words (Aksoy, 2014; AlShaikhi, 2011; Hashemi and Gowdasiaei, 2005; Hoshino, 2010); (2) a better understanding of semantic content of related words, and consequently, higher accuracy and precision of expression (Jullian, 2000), and (3) positive feedback from the learners themselves (Aksoy, 2014; Channell, 1981).

On the other hand, the studies that argue against semantic clustering report low recall rates for semantically grouped words, especially in the case of beginning learners (Erten & Tekin, 2008; Papathanasiou (2009).

The differences in target learner groups and wide variations in methodologies applied make comparison of the findings across these studies very difficult. For example, unlike earlier research which measured the amount of time participants needed to learn the target sets, in the Hashemi and Gowdasiaei (2005) study, increases in the depth and breadth of vocabulary knowledge were assessed using *The Vocabulary Knowledge Scale (VKS)* (Wesche & Paribakht, 1996). Furthermore, the study did not make a clear distinction between thematic and semantic grouping. While the 'lexical-set' condition was contrasted with 'semantically-unrelated' words, the grouping criteria included both semantic and thematic relatedness. Finally, the target words in both conditions were presented in short sentences and learners were asked to try to infer their meanings before the definitions were provided. The relatedness of target words at both semantic and topical levels and the use of a contextual framework make comparison with earlier studies problematic. The problem of classification criteria can also be observed in Papathanasiou (2009) study, in which 'semantic clusters' included cognitively based thematic word groups (e.g., *smuggling, terrorism, forgery* etc. grouped under the topic of CRIME), homonyms (e.g., *pane-pain, steak-stake* etc.) and sets based on pure sense relationships such as synonym and antonym pairs (e.g., *torment-torture, jab-punch / ebb-flow, gloom-gee*). Yet, in other papers, such as Hoshino's (2010) study, synonyms, antonyms, and hyponyms

were examined as separated categories. This lack of consistency in definition of semantic clusters makes it difficult to compare the findings of different classroom-based studies or relate them to earlier experimental research.

Learners' level of cognitive maturity may also have affected the findings. The two studies in which semantic clustering was not found to facilitate vocabulary recall involved young learners. In Erten's and Tekin's (2008) study, the participants were 60 primary school students. Research on the development of word associations in L1 showed that young children initially form phonological associations based on the rhyming properties of words. Syntactic associations appear later, followed by paradigmatic associations arising from continued language exposure and an increase in the depth of word knowledge (Entwisle, 1966). It is therefore possible that the semantic information integrated in the L1 lexical entries of child learners was not sufficient to enable them to take advantage of the categorical grouping of the target items. Papathanasiou's (2009) study involved both children and adults. However, while children were intermediate learners, adults were beginners. Adult beginners achieved significantly higher scores on the unrelated word set than on semantically grouped words in both immediate and delayed post-tests. There was no significant difference in the performance of intermediate child learners between the two conditions. Papathanasiou concluded that semantic clustering impedes L2 vocabulary at beginner level. However, due to the study design, it is difficult to tell whether the observed differences resulted from differences in learners' proficiency level or their cognitive maturity or motivation which, according to Papathanasiou, was low among child participants.

In summary, available data on the impact of semantic clustering on L2 vocabulary learning must be taken as inconclusive at best. While concerns about possible interference among semantically related items cannot be dismissed, there is also a body of evidence emerging from experimental studies, teachers' observations, and learners' reports for a positive effect of semantic grouping on the retention of novel vocabulary. Thus, more research is needed before it is understood how semantic relatedness among L2 words may affect their recognition and recall.

The position of this study is that any decision about the presentation mode of the target vocabulary should be based on data collected from L2 learners in natural conditions, using real words in the language they are trying to master. The remainder of this paper will report the results of an experiment involving a group of intermediate Japanese students of English.

2. Present study

2.1 Purpose and hypotheses

This study aimed to examine the effect that the semantic clustering of new words has on L2 vocabulary acquisition. The productive and receptive learning of semantic sets was compared to learning rates for thematic clusters and word groups based on their formal similarity. The hypothesis was that semantic grouping would be more effective than the other two conditions as it would highlight

conceptual links between lexical items, and thus facilitate the integration of new words into the mental lexicon.

A cluster of formally similar words was introduced as a control variable. Confusion between *Tuesday* and *Thursday*, which has been (e.g., Folse, 2004; Webb & Nation, 2017) commonly attributed to interference between the meanings, could easily have arisen from the formal similarities between the two words. Several studies have documented how the similarity of spoken or written forms can result in lexical confusion during both language comprehension and production (Kocic, 2008; Laufer 1981, 1988; Olsen, 1999). Therefore, it was predicted that words which shared formal properties would be more difficult to learn than the semantic or the thematic set due to confusion of similar lexical forms.

2.2 Participants

The study involved 23 Japanese college students enrolled on a 1st year general English course. The students had an intermediate level of English proficiency (approximately B1 level on the CEFR scale). The class met once a week for 90 minutes.

2.3 Procedures and materials

2.3.1 Criteria for vocabulary selection

Based on the premise that explicit vocabulary instruction should never be independent of the classroom curriculum, an effort was made to integrate the experiment into regular coursework. It was therefore important to ensure that the target words were both new to the students and related to the content of the course.

The main textbook used in the course was Craven's and Sherman's (2011) *Q: Skills for Success Listening & Speaking Level 3 (1st ed.)* published by Oxford University Press. The experiment was integrated with Unit 7 and Unit 8. In Unit 7, the students learned how to express different kinds of numerical data. Information in the textbook was expanded to include the following set of 14 words grouped under the conceptual category of *numbers*:

| | | | | | | |
|-------------------|-------------------|-----------------|--------------------|-----------------------|--------------------|----------------|
| <i>figure</i> | <i>cardinal</i> | <i>ordinal</i> | <i>Fraction</i> | <i>Numerator</i> | <i>denominator</i> | <i>decimal</i> |
| <i>statistics</i> | <i>arithmetic</i> | <i>addition</i> | <i>subtraction</i> | <i>multiplication</i> | <i>division</i> | <i>nought</i> |

The English to Japanese translation test that the students took one week before the commencement of the treatment indicated no prior knowledge of the selected words.

As can be seen from the list above, all words in the cluster are semantically related, but the connections between them are of a different type. For example, *cardinal* is a hyponym of *figure* and a co-hyponym of *ordinal*. *Figure* and *arithmetic* are functionally related. *Fraction* entails a part-whole relationship between *numerator* and *denominator*. A decision not to restrict the set membership to one type of relations was made in response to the overwhelming tendency in research on semantic clustering effects in language teaching to focus on categorical grouping and co-hyponyms. While superordinate-hyponym hierarchies within semantic fields allow the taxonomical grouping of lexical items, semantic relations are not limited to class inclusion. Lexical networks in the brain consist of many interrelated and overlapping subnetworks that involve multiple and complex semantic relations. For example, words can be related through synonymy, antonymy, or by function or entailment. In fact, there is some evidence from word association studies that suggests that for concrete words function-based relationships may be stronger than the links between hyponyms and their superordinate terms. Data from Palermo and Jenkins's (1964) free association experiment showed that both children and adults associated *table* with words such as *eat* or *food* rather than *furniture*. It was thus assumed that presenting the target vocabulary so that it closely resembles the organization of the mental lexicon would be more conducive to learning.

The topic of Unit 8 was *Cities* and the students were asked to learn the following thematically related new vocabulary.

| | | | | | | |
|----------------|--------------------|---------------------|-------------------|----------------|-----------------|--------------------|
| <i>dweller</i> | <i>Outskirts</i> | <i>crosswalk</i> | <i>landscape</i> | <i>Façade</i> | <i>gridlock</i> | <i>skyscraper</i> |
| <i>sewer</i> | <i>Megalopolis</i> | <i>municipality</i> | <i>pedestrian</i> | <i>Quarter</i> | <i>avenue</i> | <i>conurbation</i> |

Although thematic clusters can include words relating to different parts of speech, the set consisted only of nouns. It has sometimes been argued that grammatical class can affect the learnability of words. For example, some researchers argue that nouns are easier to learn than adjectives or verbs (Kersten & Earles, 2004; Phillips, 1981; Rodgers, 1969). Therefore, by restricting the target words to one part of speech, a possible confounding effect was eliminated.

Formally similar words were presented in the course with the rationale that they often cause miscomprehension in reading and are commonly confused in writing. The target items were selected from the '*Commonly confused words*' page on the Oxford Dictionaries website. The page lists 74 homophone pairs (words with the same pronunciation but different meanings and spelling). Like the other two conditions, the selected items belonged to the grammatical class of nouns. As the words were presented in isolation and belonged to the same part of speech, the learners had to rely on orthographic differences to distinguish between the meanings. The seven homophone pairs were as follows:

| | | | | | | |
|--------------|--------------|---------------|--------------|----------------|-------------|------------------|
| <i>aisle</i> | <i>chord</i> | <i>assent</i> | <i>cue</i> | <i>palate</i> | <i>pole</i> | <i>principal</i> |
| <i>isle</i> | <i>cord</i> | <i>ascent</i> | <i>queue</i> | <i>palette</i> | <i>poll</i> | <i>principle</i> |

2.3.2 Instructional treatment

The target word-sets were taught and tested separately in three 30-minute sessions. In all three conditions, the treatment consisted of two activities designed to help the students remember the meaning and form of the new words. First, the students were given a list of target words and their L2 definitions and were then instructed to write these words next to their L1 translations. The answers were checked in class and any pronunciation errors were corrected. The second task was a pair-work activity. Each student in the pair received one subset of seven words along with their Japanese translations. The words were presented in different sequences from those used in the first task to prevent a possible priming effect. The students were instructed to read Japanese words to their partner who was then expected to provide the English counterparts. If a student made a mistake, the “coaching partner” could refer to the translation-pair list and correct them. When all seven items were recalled correctly, the second student would become the “coach” and try to elicit the remaining items. Finally, the students were instructed to swap sheets so that each student could practice recall of the entire set. Samples of both activities can be found in Appendix I.

2.3.3 Post-tests

Format

Following the treatment, the students were given comprehension and production vocabulary tests. First, they were asked to provide English translations of Japanese words (a productive knowledge test). The second sheet was an English to Japanese translation task (a receptive knowledge test). The target words were presented in an order different from the sequences used in the treatment tasks. The students were given two minutes for each part of the test. The delayed post-tests administered two weeks after the treatment were in the same format; however, the words were presented in different serial positions. Samples of the tests can be found in Appendix II.

Scoring

On the receptive knowledge test, the students were awarded 1 point for each answer that matched the meaning of the target word. For example, *facade* was translated as 建物の正面 (*tatemono no shoumen*=front of the building) and 前面 (*zenmen*=a front, a frontage) and both answers were marked as correct.

3. Results

3.1 Overview of the experimental design

The experimental design comprised three independent variables: *semantic grouping*, *thematic grouping*, and *lexical grouping based on formal similarity*. The dependent variables were the students' scores on word comprehension and productive measures (i.e. *receptive* and *productive* knowledge tests). The data were collected at two points in time: immediately after the instructional treatment and after two weeks (*immediate* and *delayed* post-tests). The comprehension test results are reported first.

3.2 Students' performance on word comprehension tests

Prior to descriptive analysis, the data were checked for extreme values, normality of distribution, and homogeneity of variance. The screening for extreme values revealed one atypical case. This outlier was removed from the data set. The Shapiro-Wilk test was then used to check the normality of residuals. The results are shown in Table 1.

TABLE 1

The results of the Shapiro-Wilk normality test for word comprehension measures

| | Shapiro-Wilk | | |
|--|--------------|----|------|
| | Statistic | df | Sig. |
| Studentized Residual for RI_semantic | .631 | 22 | .000 |
| Studentized Residual for RI_thematic | .824 | 22 | .001 |
| Studentized Residual for RI_formally similar | .796 | 22 | .000 |
| Studentized Residual for RD_semantic | .829 | 22 | .001 |
| Studentized Residual for RD_thematic | .942 | 22 | .218 |
| Studentized Residual for RD_formally similar | .951 | 22 | .328 |

*RI= receptive immediate test; RD=receptive delayed test

These values indicate a violation of the normality assumption in the immediate receptive knowledge tests in all three conditions ($p < 0.05$). In the delayed receptive knowledge test, the normality assumption was met in the thematic and the formally similar word data sets ($p > 0.05$).

Mauchly's test was used to test for sphericity. The assumption of homogeneity of variance between all combinations of related groups was met for the independent variable, *method* ($W(2)=0.81$, $p=0.13$) and interaction of the two independent variables, *time*method* ($W(2)=0.94$, $p=0.55$).

Descriptive statistics are summarized in Table 2.

TABLE 2

Descriptive statistics for the immediate and delayed receptive knowledge tests

| | Mean | Std. Deviation | N |
|--|-------|----------------|----|
| Receptive immediate – semantic | 13.09 | 1.659 | 22 |
| Receptive immediate – thematic | 12.32 | 1.887 | 22 |
| Receptive immediate – formally similar | 10.55 | 1.765 | 22 |
| Receptive delayed – semantic | 10.55 | 3.687 | 22 |
| Receptive delayed – thematic | 10.68 | 1.887 | 22 |
| Receptive delayed – formally similar | 8.14 | 2.455 | 22 |

In all three conditions, the students performed better on the immediate than on the delayed post-test, which was expected according to the natural forgetting curve. In the immediate post-test, the highest scores were recorded in the semantic set condition ($M=13.09$, $SD=1.66$), followed by the thematic set ($M=12.32$; $SD=1.89$). In the delayed post-tests, the best results were obtained in the thematic set condition ($M=10.68$, $SD=1.89$), followed by the semantic set ($M=10.55$; $SD=3.69$).

The significance of the results was assessed through a two-way repeated measures ANOVA, which is considered sufficiently robust to moderate violations of the normality assumption (Glass, Peckham, & Sanders, 1972; Harwell, Rubinstein, Hayes, & Olds, 1992). There was no statistically significant interaction between the learning method and the time of testing ($F(2)=0.79$, $p=0.46$, $\eta^2=0.04$). The differences in word comprehension scores across the three conditions were found to be significant ($F(2)=20.29$, $p=0.00$, $\eta^2=0.49$). Posthoc comparisons using Bonferroni corrections showed that scores

in the *formal similarity* condition were approximately 2.48 points lower than scores in the *semantic set* method ($p=0.00$) and, on average, 2.16 points lower than scores in the *thematic set* ($p=0.00$).

There was a statistically significant difference in word comprehension regarding the time of testing ($F(1)=31.01$, $p=0.00$, $\eta^2=0.59$). Post-hoc comparisons using Bonferroni corrections showed that scores on the immediate post-test were, on average, 2.20 points higher than on the delayed post-test ($p=0.00$). The estimated marginal means of receptive post-test scores are shown in Figure 1.

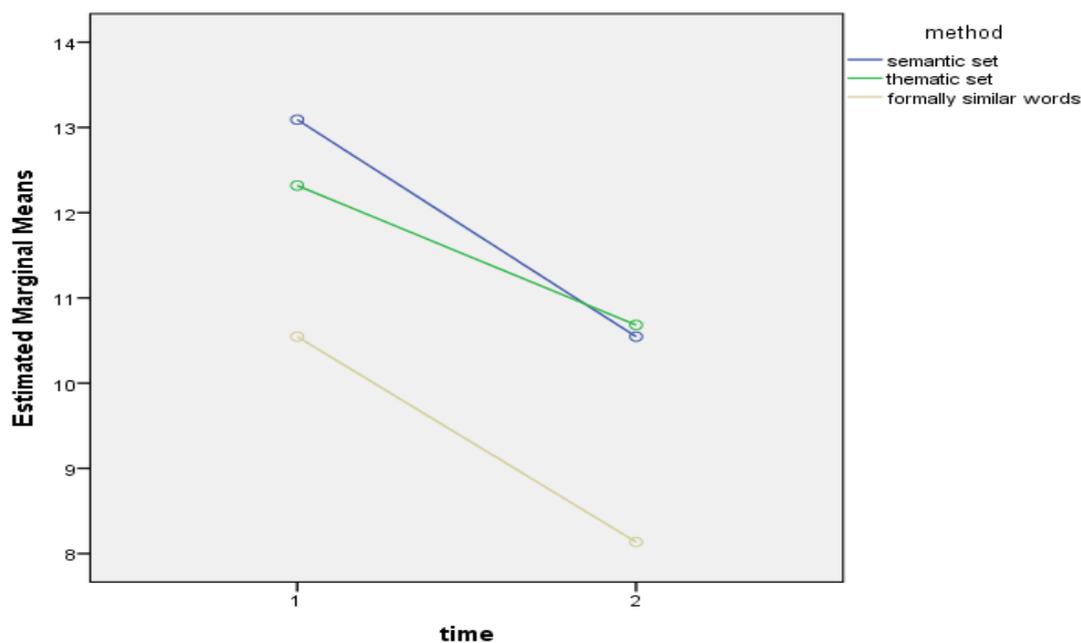


FIGURE 1

Estimated marginal means of receptive post-test scores per treatment condition

3.3 Students' performance on word recall tests (productive knowledge tests)

The productive knowledge test results were analyzed the same way as the comprehension test scores. First, the data set was examined for outliers, normality of distribution, and homogeneity of variance.

The Shapiro-Wilk test showed that the normality assumption was violated in the immediate post-test for the formally similar word set, and in the delayed post-test for the thematic set ($p<0.05$). Other measures met or approximately met the assumption of normality ($p>0.05$) (See Table 3).

TABLE 3

The results of the Shapiro-Wilk normality test for word production measures

| | Shapiro-Wilk | | |
|--|--------------|----|------|
| | Statistic | df | Sig. |
| Studentized Residual for PI_semantic | .914 | 23 | .050 |
| Studentized Residual for PI_thematic | .961 | 23 | .491 |
| Studentized Residual for PI_formally similar | .800 | 23 | .000 |
| Studentized Residual for PD_semantic | .933 | 23 | .129 |
| Studentized Residual for PD_thematic | .884 | 23 | .012 |
| Studentized Residual for PD_formally similar | .962 | 23 | .505 |

*PI= productive immediate test; PD= productive delayed test

The assumption of sphericity was met for the independent variable, *vocabulary presentation method* (Mauchly's $W(2)=0.95$, $p=0.56$), and the interaction of two independent variables, *testing time*method* ($W(2)=0.95$, $p=0.58$). No extreme values were detected.

The means and standard deviation values are shown in Table 4.

TABLE 4

Descriptive statistics for the immediate and delayed productive knowledge tests

| | Mean | Std. Deviation | N |
|---------------------------------|-------|----------------|----|
| Productive immediate – semantic | 10.48 | 2.591 | 23 |

| | | | |
|---|-------|-------|----|
| Productive immediate – thematic | 8.74 | 3.320 | 23 |
| Productive immediate – formally similar | 10.26 | 2.027 | 23 |
| Productive delayed- semantic | 4.39 | 3.381 | 23 |
| Productive delayed – thematic | 4.09 | 1.782 | 23 |
| Productive delayed – formally similar | 3.74 | 3.208 | 23 |

In the immediate productive test, the students achieved their highest scores in the semantic set condition ($M=10.48$; $SD=2.59$) followed by formally similar words ($M=10.26$; $SD=2.03$). In the delayed productive test, the students performed best in the semantic set condition ($M=4.39$; $SD=3.38$), followed by the thematic set ($M=4.09$; $SD=1.78$).

There was no statistically significant effect for an interaction between the vocabulary presentation method and the time of testing ($F(2)=1.91$, $p=0.16$, $\eta^2=0.08$) or the presentation method and the level of vocabulary recall ($F(2)=2.14$, $p=0.13$, $\eta^2=0.09$). The difference between the immediate and the delayed post-test results was found to be statistically significant ($F(1)=199.30$, $p=0.00$, $\eta^2=0.90$). Post-hoc comparisons using Bonferroni correction showed that students' scores on the immediate post-test were, on average, 5.75 points higher than on the delayed post-test ($p=0.00$). The estimated marginal means of productive post-test scores are highlighted graphically in Figure 2.

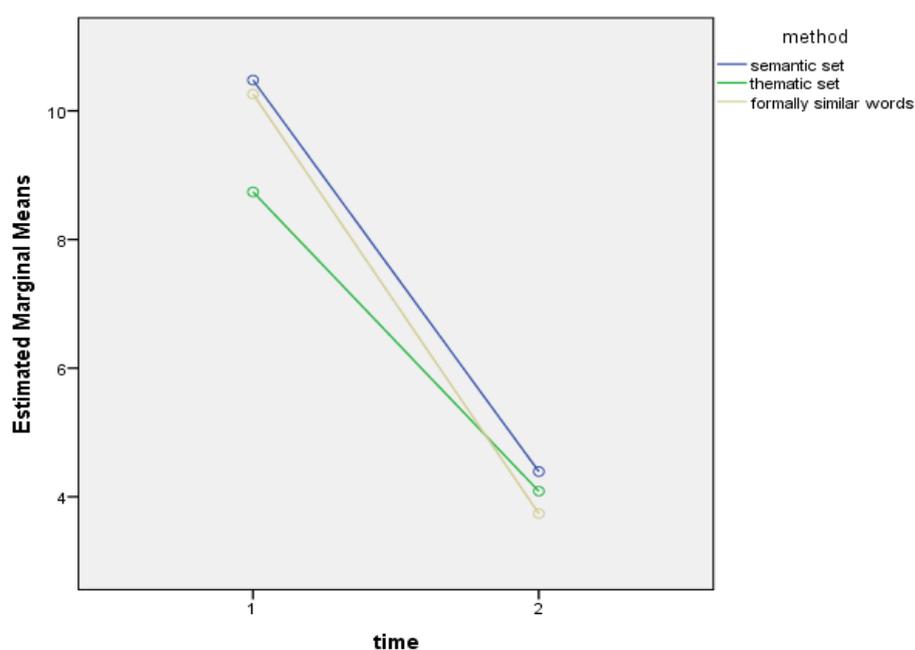


FIGURE 2

Estimated marginal means of productive post-test scores per treatment condition

This means that if a student translated 島 (shima) as *island*, they would not get any points even though *island* is a semantic equivalent of 島 (shima). Only students who recalled the target word *isle* were awarded the points.

4. Discussion

To address the limitations of previous research, the effect of using different methods of vocabulary presentation was investigated in a natural classroom environment and with real L2 words to make the task more meaningful for learners. Both receptive and productive knowledge were assessed and the immediate and long-term effects of different types of clusters were explored. Overall, the results suggest that presenting vocabulary in semantic sets is more or at least as equally effective as thematic word grouping. In the immediate vocabulary comprehension test, the students achieved the best results in the semantic set condition, while in the delayed comprehension test the difference between the mean values for the thematic set and the semantic set was minimal. While attrition was observed in all conditions, semantic clustering seemed to be mnemonically superior to other grouping methods. In both the immediate and the delayed productive knowledge tests, the highest recall rates were recorded for words from the semantic set. The findings also indicate that similarity of form may present a bigger challenge for learners than the semantic relatedness of the target words. In both the immediate and the delayed word comprehension tests, the scores for homophone pairs were lower than scores for words in either the semantic or the thematic set, and the differences were found to be statistically significant. Inferior performance was also observed in the delayed productive knowledge test, although the differences did not reach statistical significance. The only instance where the hypothesis was not confirmed was the immediate productive knowledge test, where formally similar words were recalled better than the words in the thematic set. This may be because the students, aware of the possible confusion of the forms, paid more attention to orthographic properties during the treatment stage and consequently committed fewer errors during immediate lexical recall. However, the memory traces may not have been strong enough to ensure correct word retrieval two weeks after the treatment.

Considering the small sample size and the relatively small number of words tested, the findings cannot automatically be interpreted as evidence to show that semantic clustering is the most effective way to present L2 vocabulary to learners. However, they certainly raise concerns regarding whether skepticism towards the semantic clustering of L2 vocabulary is really justified.

The discrepancies evident in the findings of different studies seem to suggest that, while semantic relatedness can cause interference, categorical membership does not automatically imply any disturbance of the learning process. It seems plausible to predict that the concurrent presentation of

conceptually related L2 words will hinder learning when there is a high level of overlap in the semantic content of the corresponding L1 lemmas. The theoretical evidence-base for this assumption can be found in Jiang's (2000) psycholinguistic model of adult L2 vocabulary acquisition in instructional settings. Building on Levelt's (1989) model of lexical representation, Jiang assumes that the entry of each word in the mental lexicon entails four kinds of information stored in two components: the *lemma component* which contains semantic and syntactic information, and the *lexeme component* in which morphological and formal information are encoded.

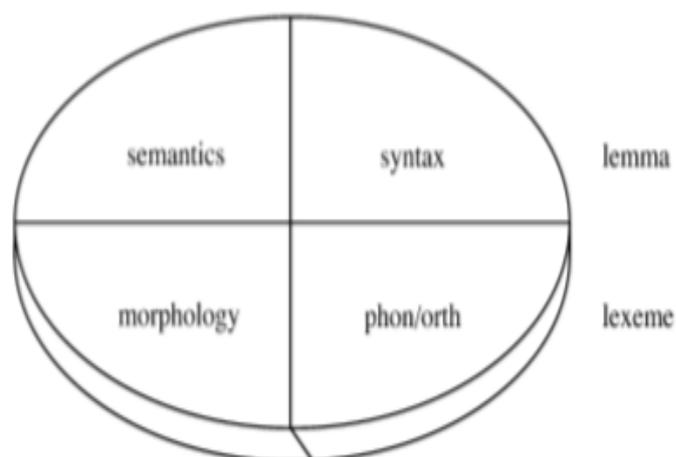


FIGURE 3

The internal structure of the lexical entry (Jiang, 2000, p. 48)

Jiang's (2000) model presumes there are three basic stages in L2 vocabulary development. The first is the *formal stage* where the form of a new L2 word is linked to a corresponding L1 word. In this stage there are no direct links between L2 words and the concept, and the semantic content of L1 words is activated during both language comprehension and production.

In the second stage, which Jiang labels the *L1 lemma mediation stage*, the semantic and syntactic information from L1 translations is copied into L2 lexical entries. This means that L1 lemma information mediates L2 word processing. L2 words are therefore linked to their conceptual representations directly, as well as through lexical association with their L1 translation counterparts. The links between L2 words and their concepts gradually grows stronger and the direct route becomes the default route.

The final stage is the *L2 integration stage* during which the semantic, syntactic and morphological information of an L2 word, typically acquired through exposure, is integrated into an L2 lexical entry. At this stage, L2 lexical entries are similar to L1 entries in terms of both representation and processing.

The three-stage development of word knowledge can be graphically presented as follows:

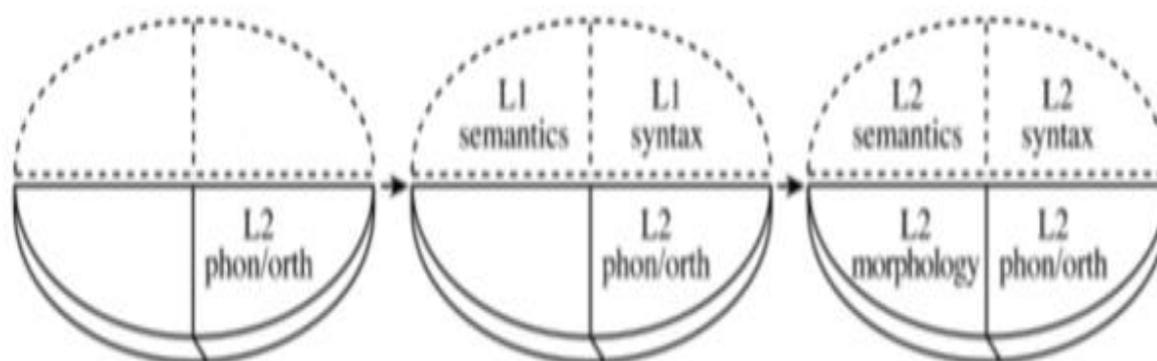


FIGURE 4

Lexical development in L2: from the formal stage to the integration stage (Jiang, 2000, p. 54)

Learners can benefit from lexical transfer when the contents of L1 and L2 lemma match. A concurrent presentation of two semantically related words should not cause more interference than the presentation of two semantically unrelated words if the corresponding L1 and L2 translations have the same semantic content. For example, learning words such as *lion* and *elephant* together should not be more difficult than learning *lion* and *table* together, assuming there are distinctive corresponding lemmas for both words in the learners' native language and the language they are trying to master.

However, interference between L2 words may occur when there is a high degree of overlap in the semantic content of the corresponding L1 lemmas. For example, unlike *hot* and *cold*, which have distinctive semantic specifications, words like *cold* and *chilly* may be confused even if they have different corresponding L2 translation. The extent of semantic overlap among the cluster members may explain some of the discrepancies observed in empirical research. For example, the semantic set in AlShaikhi's (2011) study included animals such as *raven*, *ostrich*, *hare*, *ram*, and *skunk*. Although they share categorical membership, these words have distinct semantic properties and therefore the interference effect was low. However, the set used in Erten's and Tekin's (2008) study contained words such as *hen*, *cock*, *fox*, *wolf*, *goat*, and *sheep*. Some of these words display a high level of semantic overlap as well as a high degree of visual similarity at the referent level. *Hen* and *cock* provide an illustrative example of this. The two words are conceptually similar (*hen*= female adult domestic chicken vs. *cock*= male adult domestic chicken.) Furthermore, their physical referents are visually similar. According to the dual coding theory (Paivio, 1969), referent images located in the imagery system are connected to both L1 and L2 verbal systems. These links provide an alternative lexical access which can facilitate the recognition and recall of L1 and L2 words. However, when stored images are visually similar and words are semantically close, concurrent presentation may cause interference, leading to lower recall rates for the whole cluster (Ishii, 2015, 2017).

Interference may also occur when L2 words have referents of low saliency or low codability in the learners' mother tongue (Higa, 1965). For example, for a native speaker of English, learning the

terminology for family relationships in Serbian may be very difficult, not just because the words are semantically related but because the Serbian kinship system is much more elaborate and contains many terms whose referents have low codability in English. For native speakers of Chinese, English words such as *problem* and *question* may pose difficulties as both terms have a single L1 translation (*wenti*) and the same semantic content is transferred into L2 entries at the form-meaning mapping stage (Jiang, 2002). The Spanish verb *experimentar* can mean both *to experiment* and *to experience* in English, which may lead to confusion of forms and production errors among Spanish learners of English.

Finally, interference may also be caused by the underdeveloped semantic content of L1 lemmas. The quality or completeness of semantic information in L1 entries may also explain differences in the performance of individual learners. In Tinkham's (1993, 1997) and Waring's (1997) experiments, some subjects needed approximately the same number of trials to learn the semantically related words and words in other conditions (thematic or unrelated set), while for others the difference was double or triple this number. The semantic content in an L1 lemma forms a bridge between the newly acquired L2 word and its conceptual representation. This means that, if the semantic information stored in L1 entry is insufficient, conceptual representation of the L2 term will also be underdeveloped and the possibility of semantic confusion between L2 words will be higher. For example, a learner who frequently engages in DIY carpentry is likely to have richer semantic content in lemmas for *hammer*, *saw*, *pliers* or *drill* than a learner who does not use these tools on regular basis. If both learners were asked to learn L2 words in a cluster of *household tools*, semantic interference is expected to be stronger in the case of the second learner.

In short, the difficulty of learning a set of semantically related words may depend not only on the extent of semantic overlap between the L2 lemmas but also on the distinctiveness and completeness of semantic information in the lemmas of their corresponding L1 translations.

5. Conclusion

Solid vocabulary knowledge is crucial for the development of L2 communicative ability. However, learners are faced with the enormous task of mastering potentially thousands of new words. Therefore, it is of the utmost importance that teachers do everything possible to reduce students' learning burden. Informed decisions should be made at all levels of the learning process – selection of the target vocabulary, its presentation to learners, and the choice of activities that learners engage in to commit the new words to memory.

Psycholinguistic research in L1 has shown that words in the minds of adult native speakers form complex and stable lexical networks (Postman & Keppel, 1970; Read, 1993; Wolter, 2001). Semantic fields in which words are conceptually mapped provide a basis for specification of their meanings, organization, and recall (Atchison, 1994; Öhman, 1953). Studies of word associations in L2 have shown that, as learners' proficiency increases, the structure of their mental lexicon approximates that

of native speakers (Ard & Gass, 1987; Piper & Leicester, 1980; Zareva, 2008). However, while the semantic organization of vocabulary may be indicative of the mature mental lexicon, it does not automatically imply that vocabulary is acquired in semantic sets. Among native speakers, most of the vocabulary is acquired through exposure which means that new lexical items are more likely to be encountered in thematic frames rather than semantic groups. A question that arises in this respect is whether the development of semantic structures in the mental lexicon of L2 learners with limited language exposure can be facilitated through categorical clustering of the target vocabulary. Classroom evidence offers partial but important support in favour of this hypothesis.

When learners work with real words in natural environments, the semantic sets seem to be recalled better than semantically unrelated word groups and these advantages tend to increase with the post-test time lapse. Higher level learners, who have a larger vocabulary, and older learners, who have more developed conceptual knowledge, seem to benefit more from the categorical clustering of novel vocabulary.

Semantic clustering, however, should not be applied indiscriminately. The current evidence seems to suggest that semantic clustering should be avoided when:

- 1) L2 words share the same L1 translation or when there is a high level of overlap in the semantic content of the corresponding L1 lemmas;
- 2) Semantically related L2 words have similar visual referents;
- 3) Semantically related L2 words have similar phonological or orthographic forms.

Therefore, caution is needed, although a blanket dismissal of semantic clustering approach seems to be unjustified and possibly detrimental to learners. The vocabulary of a language is not simply a long list of random items. Words are structured within memory according to certain criteria. Semantic clusters correspond to the psychological and linguistic organization of vocabulary and presenting vocabulary in semantic sets may therefore facilitate the integration of lexical items into long-term memory. Teaching small clusters of common words with clear distinctive attributes should pose little problem for learners. Furthermore, some words such as *numbers*, *days of the week*, or *months of the year* have a single underlying concept. They form close semantic sets and lend themselves naturally to categorical grouping. Allen and Vallette (1972, p. 114) give *days of the week* and *colours* as examples of words that “neither young nor older students have trouble with”, due to their concreteness. However, lexical items for which semantic boundaries are less clear, such as, for example, *dazzle*, *glow*, *glitter*, *glisten*, *twinkle* and *shimmer*, should be introduced gradually with learners’ attention drawn to the different nuances in the meaning and use of these words. Thus, any decision about whether certain words should be presented together should consider the distinctiveness of their semantic content.

The intrinsic difficulty of individual words should also be considered. Word knowledge entails a complex set of features that are not limited to inter-lexical relationships. Factors such as spoken and

written forms, morphological structure, syntactic behavior, multiple meanings, level of abstractness and idiomaticity, collocational patterns, or register restrictions can all have a significant impact on the learning burden (Laufer, 1990).

The distance between L1 and L2 is another important variable. Linguistic background determines the frequency of cognates as well as learners' perceptions of the formal properties of a word (Kocic, 2008; Laufer, 1988). Other factors such as the presence of sentential context and the number of repetitions also play a role.

Thus, the sheer complexity of word knowledge means that it may be very difficult, if not impossible, to draw blanket conclusions regarding the best way to present the target vocabulary. While it might be possible to isolate specific features of word knowledge in experimental conditions, vocabulary learning in a natural classroom is always going to be subject to the interaction of many factors. It is important to make teachers aware that organization of the target vocabulary is a subject of debate and that caution is needed. However, introducing the findings of strictly controlled laboratory experiments in vocabulary reference books without providing any background information on the context of these studies, or the methodology they use, may give instructors the erroneous impression that an impeding effect of semantic grouping is a proven fact widely acknowledged and accepted by the scholarly community.

Semantic and thematic vocabulary grouping should be considered theoretical constructs rather than pedagogical principles. Semantically related words not only share underlying concepts, they are also linked to specific cognitive frames that motivate those concepts (Fillmore & Atkins, 1992). Knowledge of these frames reflects the experiences, beliefs and practices ingrained in human existence and as such is crucial for understanding word meanings. Terms like *buyer* and *seller* are not simply two semantically related words with opposite meanings. Understanding their semantic content presumes familiarity with the *commercial transaction frame*, which also includes items such as *goods*, *cost*, and *money* (Fillmore & Atkins, 1992). This means that when learners are presented in their textbooks with a conversation between a *buyer* and a *seller*, they are introduced to a specific cognitive frame in which the two concepts have a clearly defined space. Artificial separation of the two terms based on their presumed semantic interference may be both unnatural and counterproductive.

Therefore, neither semantic nor thematic sets are intrinsically beneficial or detrimental to learning and cannot be automatically embraced or rejected. Decisions about word grouping criteria should be made cautiously and, like other aspects of teaching, selectively, taking into consideration learners' levels and objectives, the size and depth of their vocabulary knowledge, their linguistic background, and intra-lexical characteristics of the words to be learnt.

To help instructors to make better informed decisions, future research efforts should not be directed at finding a winner in the *thematic* vs. *semantic* cluster debate but at identifying the word pairs or sets that cause most disruption in language production and comprehension, so that they can be treated

properly in the classroom. To develop better materials and offer more focused instruction, teachers and writers of pedagogical materials need lists of potentially problematic items, as well as lists of clusters that have been shown to be “safe to teach” to different groups of learners at different levels of proficiency. However, until such resources are available, teachers should rely on class observations and their experience to decide how best to present the target vocabulary to a group of learners.

Although arguments in favor of semantic grouping may not be sufficient, the evidence against it is certainly not strong enough to justify the rejection of current teaching practices. Folse (2004) referred to semantic clustering as one of the seven myths of L2 vocabulary learning. It may be that *semantic cluster threat* is the myth that should be dispelled.

References:

- Aitchison, J. (1994). *Words in the mind: An introduction to the mental lexicon*. Oxford: Blackwell.
- Aksoy, F. (2014). A controversy in presenting new vocabulary in an EFL class: semantically related sets (SR), semantically unrelated sets (SU), thematically related sets (TR). *Journal of Foreign Language Teaching and Applied Linguistics*. URL:<http://oaji.net/articles/2016/3124-1458500649.pdf> (Retrieved: January 20, 2018).
- Allen, E.D. & Vallette, R. M. (1972). *Modern language classroom techniques: A handbook*. New York: Harcourt Brace Jovanovich.
- AlShaikhi, A. Z. (2011). *The effects of semantic and thematic categorization of vocabulary on Arabic-speaking EFL learners*. (Unpublished Master thesis). Colorado State University, U.S.
- Ard, J. & Gass, S. (1987). Lexical constraints on syntactic acquisition. *Studies in Second Language Acquisition*, 9(2), 235-252
- Boers, F., Piquer, A., Stengers, H., & Eyckmans, J. (2009). Does pictorial elucidation foster recollection of figurative idioms? *Language Teaching Research*, 13(4), 367–388.
- Bousfield, W.A. & Cohen, B.H. (1955). The effects of reinforcement on the occurrence of clustering in the recall of randomly arranged associates. *The Journal of General Psychology*, 36(11), 67-81.
- Channell, J. (1981). Applying semantic theory to vocabulary teaching. *ELT Journal*, 35(2), 115-122.
- Cofer, C. N., Bruce, D.R., & Reicher, G.M. (1966). Clustering in free recall as a function of certain methodological variations. *Journal of Experimental Psychology*, 71(6), 858-866.
- Commonly confused words*. | Oxford dictionaries. <https://en.oxforddictionaries.com/usage/commonly-confused-words> (Retrieved: November 3, 2017).
- Craven, M. & Sherman, K. D. (2011). *Q: Skills for success: Listening & speaking (Level 3) (1st ed.)*. Oxford: Oxford University Press.
- Craik, F.I.M. & Lockhart, R.S. (1972). *Levels of processing: A framework for memory research*. *Journal of Verbal Learning & Verbal Behavior*, 11(6), 671-684.
- Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Cruse, D. A. (1986). *Lexical semantics*. Cambridge: Cambridge University Press.
- Davies, M.J. (2012). A comparison of the memorization and recall of vocabulary in semantic and thematic groups. *Vocabulary Education & Research Bulletin*, 1(1), 3-4.
- Dunbar, S. (1992). Developing vocabulary by integrating language and context. *TESL Canada Journal*, 9(2), 73-93.
- Entwisle, D.R. (1966). *Word associations of young children*. Baltimore: Johns Hopkins University Press.
- Erten, I. H. & Tekin, M. (2008). Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System*, 36(3), 407-422.
- Fillmore, C. & Atkins, B. (1992). Toward a frame-based lexicon: The semantics of RISK and its neighbors. In A. Lehrer & E. F. Kittay (Eds.), *Frames, fields, and contrasts: New essays in semantics and lexical organization* (pp. 75-102). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Finkbeiner, M. & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics*, 24(3), 369–383.
- Folse, K. (2004). *Vocabulary myths: Applying second language research to classroom teaching*. Ann Arbor: The University of Michigan Press.
- Gairns, R. & Redman, S. (1986). *Working with words*. Cambridge: Cambridge University Press.
- Glass, G.V., Peckham, P.D., & Sanders, J.R. (1972). Consequences of failure to meet assumptions underlying fixed effects analyses of variance and covariance. *Review of Educational Research* 42(3), 237-288.
- Haberlandt, K. (1994). *Cognitive psychology*. Boston: Allyn & Bacon.
- Harwell, M.R., Rubinstein, E.N., Hayes, W.S., & Olds, C.C. (1992). Summarizing Monte Carlo result in methodological research: the one- and two-factor fixed effects ANOVA cases. *Journal of Educational and Behavioral Statistics* 17(4), 315-339.
- Hashemi, M.R., & Gowdasiaei, F. (2005). An attribute-treatment interaction study: Lexical-set versus semantically-unrelated vocabulary instruction. *RELC Journal*, 36(3), 341-361.
- Higa, M. (1963). Interference effects of intralist word relationships in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, 2(2), 170-175.
- Higa, M. (1965). The psycholinguistic concept of “difficulty” and the teaching of foreign language vocabulary. *Language Learning*, 15(3-4), 167-179.
- Hoshino, Y. (2010). The categorical facilitation effects on L2 vocabulary learning in a classroom setting. *RELC Journal* 41(3), 301-312.
- Ishii, T. (2015). Semantic connection or visual connection: Investigating the real source of confusion. *Language Teaching Research*, 19(6), 712-722.
- Ishii, T. (2017). The impact of semantic clustering on the learning of abstract words. *Vocabulary Learning and Instruction*, 6(1), 21-31.
- Ishii, T. & Maruyama, Y. (2009). How mythical are ‘Vocabulary Myths’ among Japanese learners of English? *The Journal of Rikkyo University Language Center*, 21, 17-23.

- Jenkins, J.J. & Russell, W.A. (1952). Associative clustering during recall. *Journal of Abnormal & Social Psychology*, 47(4), 818-821.
- Jiang, N. (2000). Lexical representation and development in L2. *Applied Linguistics*, 21(1), 47-77.
- Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition* 24(4), 613-637.
- Jullian, P. (2000). Creating word-meaning awareness. *ELT Journal* 54(1), 37-46.
- Kersten, A.W., & Earles, J.L. (2004). Semantic context influences memory for verbs more than memory for nouns. *Memory and Cognition*, 32(2), 198-211.
- Kocic, A. (2008). The problem of synforms (similar lexical forms). *Facta Universitatis (Linguistics and Literature)* 6(1), 51-69.
- Laufer, B. (1981). A problem in vocabulary learning – synophones. *ELT Journal* 35(3), 294-300.
- Laufer, B. (1988). 'Sequence' and 'order' in the development of L2 lexis. *Applied Linguistics*, 11(3), 281-296.
- Laufer, B. (1990). Why are some words more difficult than others? Some intralexical factors that affect the learning of words. *International Review of Applied Linguistics in Language Teaching*, 28(4), 293-307.
- Levelt, W.J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: Bradford.
- Little, D., Singleton, D., & Silvius, W. (1984). *Learning second languages in Ireland: Experience, attitude and needs*. Dublin: Trinity College, Centre for Language and Communication Studies. (ERIC Document Reproduction Service No. ED 246670. <https://files.eric.ed.gov/fulltext/ED246670.pdf> (Retrieved: April 26, 2018).
- Marzano, R.J. & Marzano, J.S. (1988). *A cluster approach to elementary vocabulary instruction*. Newark, DE: International Reading Association.
- McGeoch, J.A. & McDonald, W.T. (1931). Meaningful relation and retroactive inhibition. *American Journal of Psychology*, 43(4), 579-588.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Neuner, G. (1992). The role of experience in a content-and-comprehension oriented approach to learning a foreign language. Arnold, P. J. & Bejoint, H. (Eds.) *Vocabulary and applied linguistics* (pp. 156-166). London: Palgrave Macmillan.
- Öhman, S. (1953). Theories of linguistic field. *Word*, 9, 123-134.
- Olsen, S. (1999). Errors and compensatory strategies: a study of grammar and vocabulary in texts written by Norwegian learners of English. *System*, 27(2), 191-205.
- Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, 76(3), 241-263.
- Palermo, D.S. & Jenkins, J.J. (1964). *Word association norms: Grade school through college*. Minneapolis: University of Minnesota Press.
- Papathanasiou, E. (2009). An investigation of ways of presenting vocabulary. *ELT Journal*, 63(2), 313-322.

- Phillips, T.A. (1981). *Difficulties in foreign language vocabulary learning and a study of some of the factors thought to be influential* (master's thesis). Birkbeck College, University of London.
- Piper, T.H. & Leicester, P.F. (1980). Word association behavior as an indicator of English language proficiency. Retrieved December 4, 2002, from Educational Resources Information Center (ERIC) documents [online]. (ED 227651).
- Postman, L. & Keppel, G. (1970). *Norms of word associations*. New York: Academic Press.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10(3), 355-371.
- Rodgers, T.S. (1969). On measuring vocabulary difficulty: an analysis of item variables in learning Russian-English vocabulary pairs. *International Review of Applied Linguistics* 7(4), 327-343.
- Seal, B. D. (1991). *Vocabulary learning and teaching*. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (pp. 296-311). Boston: Heinle & Heinle.
- Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System*, 21(3), 371-380.
- Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research*, 13(2), 138-163.
- Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review*, 69(4), 344-354.
- Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, 64(1), 49-60.
- Waring, R. (1997). The negative aspects of learning words in semantic sets: A replication. *System*, 25(2), 261-274.
- Webb, S. & Nation, P. (2017). *How vocabulary is learned*. Oxford: Oxford University Press.
- Wesche, M. & Paribakht, T.S. (1996). Assessing second language vocabulary: depth versus breadth. *The Canadian Modern Language Review* 53(1), 11-40.
- Williams, H. (2003). The effects of memorizing Thai vocabulary in semantic and thematic sets: A replication. *The ORTESOL Journal*, 21, 41-65.
- Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition*, 23(1), 41-69.
- Zareva, A. (2007). Structure of the second language mental lexicon: how does it compare to native speakers' lexical organization? *Second Language Research*, 23(2): 123-153.

APPENDIX I

Samples of the teaching materials

Task I

Instructions: Read the definitions below and write the words in bold next to their Japanese translations.

1. **figure** = an amount or value expressed in numbers
2. **cardinal** = a number such as 1,2,3, used to show quantity rather than order
3. **ordinal**= a number that refers to the position of something in a series, for example, 'first', 'second'
etc.
4. **fraction** = a number usually expressed in the form a/b

1. 足し算 _____
2. 分子 _____
3. 算数 _____
4. 順序数 _____

.....
Task II

Student A

Part One

Instructions: Read the following Japanese words to your partner and ask him/ her to give you corresponding words in English. Check your partner's answers against the model answers below.

| Japanese words | <i>Model Answers</i> |
|----------------|----------------------|
| 算数 | Arithmetic |
| 統計 | Statistics |
| 足し算 | Addition |
| 順序数 | Ordinal |
| 分子 | Numerator |
| ゼロ | Nought |
| 数字 | Figure |

Part Two

Instructions: Listen to your partner and provide English translations of the words you hear.

SWITCH!

APPENDIX II

Test samples

NUMBERS

Task I

Instructions: Translate the following words into English:

1. ゼロ _____
2. 引き算 _____
3. 掛け算 _____
4. 分母 _____
5. 割り算 _____

.....

Task II

Instructions: Translate the following words into Japanese:

1. addition _____
2. fraction _____
3. arithmetic _____
4. cardinal _____
5. subtraction _____