Effects of semantic clustering and repetition on incidental vocabulary learning

Mercedes Pérez-Serrano[1], Marta Nogueroles-López[2], Jon Andoni Duñabeitia-Landaburu [3,4]*

[1]Departamento de Lengua Española y Teoría de la Literatura, Universidad Complutense de Madrid, Madrid, Spain
[2]Departamento de Filología, Comunicación y Documentación, Universidad de Alcalá, Alcalá de Henares, Spain
[3]Centro de Investigación Nebrija en Cognición (CINC), Universidad Nebrija, Madrid, Spain
[4]Department of Languages and Culture, The Arctic University of Norway, Tromsø, Norway

* Correspondence:
Corresponding Author
jdunabeitia@nebrija.es (J.A.D.)

Abstract
The present study intended to investigate, first, the impact of semantic clustering on the recall and recognition of incidentally learned words in a new language, and second, how the interaction between semantic clustering and frequency of occurrence may modulate learning. To that end, Spanish university students watched an intentionally created video which contained Spanish target words that were either semantically related to others of the set, or not semantically linked at all. Furthermore, frequency of appearance changed among target words (1 | 4 | 8). All these words were paired with pseudowords that appeared as on-screen text during the videos. Participants were completely naive to the phases and the procedure of the experiment. After viewing the video, participants completed a recall test and a recognition test. Results showed that words presented in semantically unrelated categories were better recalled and better recognized than those presented in semantic clusters, especially when the words were presented more often.

1   Introduction

Incidental vocabulary learning occurs when learners are not forewarned that a vocabulary test will follow (Hulstijn, 2001), and thus such learning takes place as the by-product of a meaning-focused task (Chen & Truscott, 2010; Ellis, 1999; Wesche & Paribakht, 1996); whereas intentional vocabulary learning takes place through exercises and activities (such as flash cards, fill in the blanks, or matching exercises) that are designed to explicitly focus students on learning given words (Webb, 2020b). It is generally acknowledged that incidental learning is an essential component of second language (L2) vocabulary development (Webb & Nation, 2017). For a word to be learnt, either incidental or intentionally, a single exposure to the target item is not always sufficient. In fact, the frequency of exposure to a word has been proved to promote vocabulary learning with audiovisual input (Malone, 2018; Peters et al., 2016; Peters & Webb, 2018), so that the more times users are exposed to a word, the more likely they are to recognize it and remember it.

Semantic clustering on incidental vocabulary learning

Another variable that, together with frequency of occurrence, may have an impact on vocabulary learning is the way new vocabulary is presented. In the field of foreign language teaching, vocabulary has been traditionally taught in semantic groups consisting of words that are arranged together according to the semantic field and the syntactic word class they belong to (e.g., Marzano & Marzano, 1988). Despite this common practice in second language pedagogy, previous studies do not find agreement on whether this type of teaching causes interference or not in the learning of the target words due to the semantic similarity of such linguistic units. The present study focuses on the effects of semantic clustering on vocabulary learning as well as on the interaction between semantic clustering and frequency of occurrence, as such aspects have not been thoroughly explored in incidental learning situations.

The issue of incidental vocabulary learning in a second language has received considerable critical attention. While most research has focused on reading (Horst et al., 1998; Pellicer & Schmitt, 2010; Sánchez-Gutiérrez et al., 2019) and some on listening (Uchihara et al., 2021; van Zeeland & Schmitt, 2013), recent evidence suggests that vocabulary can also be learnt incidentally through viewing short videoclips (Montero Pérez et al., 2014) and single full length television programs (Peters & Webb, 2018). Videos are considered an appropriate source of vocabulary learning since they provide both ecologically valid learning situations, as well as repeated encounters with low-frequency words through a relatively small amount of viewing (Webb & Rodgers, 2009). Additionally, research has shown that vocabulary is better learned when learners are exposed to simultaneous channels of information, namely, the auditory discourse, the written captions, and the visual images (Bisson et al., 2013; Rodgers, 2018; Pérez Serrano et al., 2021).

There is robust evidence that words must be encountered many times in the input -be it written, aural or bimodal- in order for considerable gains to happen. Thus, frequency of occurrence is a critical factor that modulates incidental vocabulary learning, also with audiovisual input (Peters et al., 2016; Peters & Webb, 2018). As shown by Peters et al. (2016), there is a positive correlation between vocabulary growth and frequency of occurrence. In this line, on-screen text aids have also been shown to provide benefits for learners in terms of vocabulary learning depending on their frequency of appearance (Malone, 2018). Apart from the frequency of occurrence of the target items in the input, existing research recognizes the crucial role played by other critical factors in vocabulary growth. Vocabulary size (Montero Pérez, 2020; Peters & Webb, 2018) and working memory skills (Montero Pérez, 2020) are some of the learner-related factors with an influence on learning, while word relevance, cognateness (Peters & Webb, 2018) and the orthotactic pattern of the words (Pérez Serrano et al., 2021) are some of the word-related factors that have been identified so far. However, there is one unexplored factor that is expected to modulate incidental vocabulary development and that has been consistently manipulated in educational contexts: the semantic relation between the target words presented in the input.

In L2 instructed settings, it is common to find that vocabulary is presented semantically clustered. Following Tinkham (1997), a semantic cluster contains words pertaining to the same semantic field (e.g., arm, leg, foot, hand, shoulder, etc.) which fall under the superordinate concept (e.g., parts of the body). Words semantically clustered also present syntactic similarities, since they all belong to the same syntactic word class. Importantly, this practice has originated some scientific debate around whether presenting words in semantic clusters promotes or hinders vocabulary learning, since empirical evidence remains inconclusive (see Bürki et al., 2020, for review). Behind those who defend the detrimental effects of semantic clustering on vocabulary learning, one can find the Interference theory and Distinctiveness theory. Interference theory (Baddeley, 1990; Higa, 1963) hypothesizes that similarity between language items to be learned increases the difficulty of learning and remembering such items (see Crowder, 1976). In line with this and focusing on the differences rather than on the similarities, the Distinctive hypothesis essentially states that differentiated

Semantic clustering on incidental vocabulary learning

language items are easier to learn (Hunt & Elliot, 1980; Hunt & Mitchell, 1982). In other words, according to these theories, semantic clustering impedes vocabulary learning because it causes interference between similar meanings of related words and creates competing memory traces. According to this, lexical items should then be presented in a non-related way, allowing learners to differentiate the target items. Based on these two theories, many researchers argue against semantic clustering as they claim that accessing semantically related words at the same time makes learning more difficult (e.g., Baddeley, 1997; Barcroft, 2015; Folse, 2004; Korochkina et al., 2021; Nation, 2000, 2013; Nation & Webb, 2011; Schmitt, 2007, 2010; Waring, 1997; Wilcox & Medina, 2013). In sharp contrast, other researchers support that semantic clustering facilitates vocabulary learning (e.g., Grandy, 1992; Hashemi & Gowdasiaei, 2005; Haycraft, 1993; Hoshino, 2010; Seal, 1991; Stoller & Grabe, 1995; Wharton & Race, 1999), based on arguments from the Semantic field theory (Lehrer, 1974). This theory sustains the view that words with related meanings are closer in our mental lexicon. Thus, in order to mimic how vocabulary is organized in our minds, it should be presented in semantic sets (Nation, 2000). This way, such authors claim that it would be the most natural and logical way to present vocabulary in instructed settings. Additionally, as Nakata & Suzuki (2018) point out, the presumed difficulty posed by competing similar items in related sets might be beneficial in the sense that difficult items attract more attention and engagement from learners (Finkbeiner & Nicol, 2003).

Among the studies which found negative effects of semantic clustering on vocabulary learning, Tinkham (1993) compared the number of trials needed to recall target pseudowords previously presented with words in a semantically related and unrelated fashion. He found that significantly more trials were needed in the case of related words. Similarly, Wilcox & Medina (2013) found that semantically clustered (and not phonologically similar) words were more difficult to learn. Papathanasiou (2009) found a difference modulated by participants' age and level of proficiency: semantic clustering would hinder learning for beginner adults, but not for children with an intermediate level. More recently, Korochkina et al. (2021) examined whether intended learning of new words was facilitated by presenting them together with words from the same semantic category, as opposed to with words from different semantic categories. Their experiment consisted of two sessions conducted on two consecutive days in which participants learned novel word-forms in German for familiar concepts, and were tested only in the second session through picture naming, translation and picture-word interference tasks. Items were intentionally displayed in groups of four belonging all four to either a related or unrelated context. As this experiment was framed in an explicit or intended learning scenario, participants were encouraged to practice and repeat words during the learning phase. The authors concluded that teaching new words together with other words from the same semantic category resulted in poorer accuracy in the learning phase and slower lexical access in the recall tasks. In this same line, Nakata & Suzuki (2018) studied the effects of semantic clustering on L2 vocabulary learning as well as the interaction between spacing and the semantic clustering of isolated lexical items in intentional learning. They tested whether spacing between target words enhanced the learning of semantically related words by reducing interference. While they did not find any significant differences between semantically related and unrelated sets in translation accuracy on the posttests, semantically related sets caused more within-set errors than did unrelated sets. They also found that in the massed group, semantically related items resulted in a larger number of interference errors than unrelated items on both the immediate and delayed posttests. As a consequence, their results also support the widely held view that semantic clustering should be avoided, at least when presented in massed sets.

In spite of the existing evidence, and probably as a consequence of generalized practices found in textbooks, there is still uncertainty whether semantic clustering is beneficial or not in educational contexts. While there is a growing body of research on the effects of semantic clustering, such studies exclusively focus on intentional learning scenarios. However, it is widely accepted that incidental

vocabulary learning is central to lexical development (Webb, 2020a). The present study aims to address this issue by manipulating the semantic relatedness of new target words presented while participants viewed a subtitled video. Furthermore, considering that preceding research has shown that incidental vocabulary learning is modulated by the frequency of appearance of the items (Pigada & Schmitt, 2006; Rott, 1999; Waring & Tataki, 2003), in the current study, number of exposures was also manipulated. Interestingly, the interaction between both factors, namely, semantic clustering and frequency of occurrence, in incidental learning situations has not yet been thoroughly investigated. Thus, we designed a study to explore whether semantically unrelated words that are incidentally learned require less frequent occurrences to be learned.

2 Materials and methods

2.1 Participants

One hundred and two (70 women; mean age= 19) participants, aged between 18 and 22, were involved in the study. All of them were university students majoring in Education at University of Alcala and in Applied Languages and Linguistics at Complutense University of Madrid, and thus living in Spain, with Spanish as their L1. Before the experimental session, all participants gave their informed consent in accordance with guidelines approved by the Ethics and Research Committees of the Nebrija University. The experiment was implemented during synchronous online sessions with the researchers being present for problem solving if required. Participants were asked in advance to use an individual PC (rather than a tablet or a smartphone) and a headset.

2.2 Design and stimuli

For the experiment, we used a repeated-measures design with Type of Item (semantically clustered | non-semantically clustered) and Number of Exposures (1 | 4 | 8) as within-subject factors. Following Pérez Serrano et al. (2021), we intentionally created one video for this purpose, ensuring full control over the factors that could influence the results, such as the order of appearance of the targets, their particular linguistic features and the number of exposures to them. The soundtrack of the video consisted of 52 sentences in Spanish containing a total of 12 target words, as well as an opening and a closing sentence with no targets embedded. Each one of the 52 sentences contained a single mention to one of the Spanish targets. Half of the targets were semantically related, a set consisting of six coordinates as they all fell under the category of *fruits* (i.e., *grape, apple, pear, orange, mango, strawberry*). The use of coordinates, rather than synonyms or antonyms, is justified because previous studies have found significant effects of semantic relatedness using this type of semantically related targets in their designs (Nakata & Suzuki, 2018; Finkbeiner & Nicol, 2003; Tinkham, 1993, 1997). The other half were not semantically related at all, each one of them corresponding to independent categories such as tools, means of transportation, parts of the body, cutlery, clothes, tools, vehicles and animals (i.e., *finger, fork, skirt, hammer, motorbike, rabbit*). All the targets were controlled for L1-related factors, such as frequency of use in Spanish (zip. mean 3.79; range 3.27 to 4.30), part of speech (all of them are nouns) and other semantic properties such as concreteness (mean 5.86; range 5.11 to 6.36)[1]. In each category, semantically related and unrelated, there were two items that appeared eight times, two that appeared four times and two that appeared only once. Each Spanish target word was paired with an invented two or three syllable pseudoword, that is, strings that do not exist in participants' language, but that contain letter combinations that are orthotactically and phonotactically plausible (see Table 1). The invented words were created using Wuggy (Keuleers & Brysbaert, 2010).

---

[1] The lexical properties of the words were taken from Espal (Duchon et al., 2013).

All of them were pronounceable in the L1 and contained legal combinations of letters in Spanish because previous literature has demonstrated that orthotactic markedness may affect the learning burden (Pérez Serrano et al., 2021).

The sentences contained objective information on the targets, such as "*The first motorcycle reached a speed of eighteen kilometers per hour*". The texts were transformed into speech using an online software and the resulting audio clips were included in the videos using commercial video editing software. The verbal content was accompanied by supporting related visual information depicting the target words mentioned, which appeared on screen during the utterance of each sentence (see also Pérez Serrano et al., 2021). The captions of the video appeared in the lower part of the screen and displayed the entire utterance, replacing the target words with the corresponding pseudowords in capital letters ("*The first INA reached a speed of eighteen kilometers per hour*"). The verbal input, except from the pseudowords, was in Spanish. Two different versions of the video clip were created in order to ensure that the place of appearance of the targets did not affect the results and thus to avoid cumulative semantic interference being a confounding variable. To do so, the order of appearance of the sentences in the audio files was randomized. The length of the videos was 6 minutes and 9 seconds. Each participant was randomly assigned one version of the video.

2.3   Procedure

The experimental session lasted about 20 minutes in total and all the participants used an individual PC with a headset to complete the task. Materials were presented using Gorilla (Anwyl-Irvine et al., 2019). After the consent form and a short demographic questionnaire, participants watched their assigned version of the captioned video. No heads-up was given on the purpose of the experiment, the format of the videos or the tests that would come after. They were only instructed to pay attention to the screen, which ensured a true incidental learning situation where the association between the aural input (e.g., *motorbike*) and the on-screen text (e.g., the pseudoword *INA*) could easily occur. Once participants had watched the video, they completed a simple timed N-back filler task that was created to divert their attention from the content of the videos. At the end of it, participants took two tests aimed at assessing their gained productive and receptive knowledge of the correspondence between the new words (namely, the pseudowords) accidentally presented in the captions and the Spanish words played in the soundtrack (namely, the names of the fruits and the other entities). First, a recall task was presented. Students were given all the images previously displayed in the video one by one, and they were asked to type the pseudoword they thought that corresponded to each image. This task was followed by a recognition test, in which participants were provided with a pseudoword and twelve different images of the entities presented in the videoclips. The task consisted of choosing among all the images the one corresponding to the given pseudoword. This was done once for each of the twelve critical pseudowords in a random order for each participant.

3   Results

A series of analysis of variance (ANOVAs) with Relatedness (semantically related | unrelated) and Number of Exposures (1 | 4 | 8) as within-subject factors were run to answer the research questions. The descriptive statistics of the different tasks split per condition are shown in Table 2 and Figure 1.

- Insert Figure 1 and Table 1 around here -

3.1   Recall test

We first considered participants' success or failure in typing the exact string of letters of the pseudoword. The ANOVA showed a main effect of Number of Exposures (F(2,202)= 56.29, p<.001, $\eta^2_{partial}$ = .358), signaling that learning gains increased as a function of repetitions, and an interaction between Number of Exposures and Relatedness (F(2,202)= 11.92, p<.001, $\eta^2_{partial}$ = .106). The main effect of Relatedness was not found to be significant (F(1,101)= 2.07, p=.15, $\eta^2_{partial}$ = .020). Holm-corrected pairwise comparisons indicated that the recall for items in the unrelated set increased with repetitions (1 vs. 4: t(101)=5.72, p<.001; 4 vs. 8: t(101)=4.58, p<.001), while this effect was much smaller for items in the related set (1 vs. 4: t(101)=4.19, p<.001; 4 vs. 8: t(101)=1.49, p=0.419; see Figure 1). When exploring the differences between the two Relatedness levels for each of the levels of Number of Exposures, results showed that the only significant differences were found when items were repeated 8 times (t(101)=3.85, p<.001), with the other pairwise comparisons resulting not significant (ps>.10; see Figure 1)[2].

Since precise recall cannot be expected in an incidental learning task like the present one, we carried out further analysis on these data by computing the Levenshtein Distance (LD). LD is a string metric that measures the difference between two sequences, calculated by counting the number of edits (insertions, deletions or substitutions) required to change one string into the other. In this case, LD reveals a more accurate measure of recall, allowing us an estimation of the degree of similarity between the typed sequence and the expected one, with lower values meaning better performance. An ANOVA with the same factors and levels was carried out, and results revealed main effects of Number of Exposures (F(2,202)=20.12, p<.001, η2partial = .166) and Relatedness (F (1,101)=4.28, p=0.041, η2partial = .041), and an interaction between the two factors (F(2,202)=13.32, p<.001, η2partial = .116). Post-hoc tests indicated that the number of edits needed to convert the typed string into the target decreased with repetitions, but this decrease was more marked for items in the unrelated set (unrelated set: 1 vs. 4: t(101)=1.38, p=0.682; 4 vs. 8: t(101)=5.91, p<.001; related set: 1 vs. 4: t(101)=3.93, p=0.002; 4 vs. 8: t(101)=0.84, p=0.845). Similarly, the difference between Relatedness levels was only significant when items were repeated 8 times (t(101)=4.81, p<.001), and no significant pairwise contrasts were found when items were presented once or 4 time (ps>.52).

- Insert Table 2 around here -

3.2   Recognition test

Recognition data were analyzed following the same design. Participants showed an overall good level of accuracy in their recognition performance, since the mean percentage of errors was 47.5% in a task in which they had to choose among 12 possible candidates (see Figure 1 and Table 2). The

---

[2] The same set of analysis was repeated using linear mixed effect modelling, including Relatedness and Number of Exposures (as a scaled and centered numerical variable) and their interaction as fixed effects and Participants as a random factor. Fixed effect omnibus tests are presented next, with the Satterthwaite correction method for degrees of freedom. When the main accuracy in the recall task was used as a dependent variable, the results replicated those obtained in the ANOVA, with a significant Number of Exposures effect (F(1,1119)=107.24, p<.001) and a significant Number of Exposures * Relatedness interaction (F(1,1119)=16.99, p<.001). The simple effects also showed that the effect of the Number of Exposures was larger for the unrelated (F(1,1119)=104.8, p<.001) than for the related set (F(1,1119)=19.4, p<.001). Parallel results were obtained when considering the Levenshtein Distance as the dependent variable from the recall test, with a significant Number of Exposures effect (F(1,1119)=36.79, p<.001) and a significant Number of Exposures * Relatedness interaction (F(1,1119)=11.05, p<.001). The simple effects also showed that the effect of the Number of Exposures was larger for the unrelated (F(1,1119)=44.08, p<.001) than for the related set (F(1,1119)=3.76, p=.053). The analysis of the data from the recognition test using linear mixed modelling showed a significant effect of Number of Exposures (F(1,1134)=96.80, p<.001) and of Relatedness (F(1,1134)=10.70, p=.001), with no interaction between the factors (F<1 and p>.45), thus replicating the same results obtained in the ANOVA.

ANOVA showed main effects of Number of Exposures ($F(2,202)=38.74$, $p<.001$, $\eta^2_{partial} = .277$) and Relatedness ($F(1,101)=10.84$, $p=.001$, $\eta^2_{partial} = .097$), but no interaction between the two factors ($F(2,202)=2.03$, $p=.134$, $\eta^2_{partial} = .020$)). Elements presented in unrelated categories were better recognized than those presented in related categories, and recognition improved as a function of repetitions.

4   Discussion

The current study provides the first known piece of evidence on the role of semantic clustering in incidental vocabulary learning. Previous studies evaluating the effects of semantic clustering on deliberate or intentional learning scenarios have reported inconsistent results on whether this way of presenting lexical items is beneficial for learners or not. Considering that the precise role of semantic clustering has not been explored in incidental learning scenarios, and that the manner in which this effect could interact with a repeatedly tested factor like the number of encounters with the target items has not been investigated to date, the present study was set to fill this gap.

The main question at test in this study sought to determine the effects of semantic clustering on incidental learning of words. In line with other studies like those by Finkbeiner & Nicol (2003), Papathanasiou (2009), and Tinkham (1993), our data consistently indicates that words presented in semantically related sets are more difficult to learn. Results from the recall test, considering both the success or failure in typing the exact learnt string as well as an index of similarity between the recalled string and the target one (e.g., the Levensthein distance), showed that learning is more likely to occur when words are presented with no semantic competitors, that is, in unrelated sets, especially when maximizing the exposure to the items to be learnt. Likewise, the analysis performed on the recognition data showed that learners are more successful in associating the correct name of the elements with their pictorial representation when the items had been presented in semantically unrelated sets during the learning phase. Such results are in line with Nakata & Suzuki's study (2018), as they found that semantically related words resulted in a larger number of within-set errors than unrelated items, and thus, together these data support the view that semantic clustering should be avoided during learning. Our results are also consistent with the findings by Korochkina et al. (2021) who found that, as compared to the unrelated words, the categorically related items resulted in poorer naming accuracy in the learning phase, slower response latencies at the immediate recall tasks and greater semantic interference in the picture-word interference task. However, unlike Korochkina et al., in the current study the novel words were not presented blocked by group and learning context, providing additional evidence on learning scenarios that are closer to what a learner could experience in a real context. While in their case items belonging to the related group were presented in one block and those belonging to the unrelated context in another, in the present study, items were intermixed in the exposure and test phases.

Consistent with the literature on the role of the frequency of occurrence in incidental vocabulary learning, these data also demonstrated an effect of the number of exposures on the learning of the new words, showing that the more times a word is encountered, the more likely it is to be both recognized and recalled. Indeed, our results support preceding studies which confirm that frequency is closely tied to incidental vocabulary learning not only after exposure to written (Pellicer-Sánchez & Schmitt, 2010; Webb, 2007) and aural input (Brown et al., 2008; van Zeeland & Schmitt, 2013), but also with audiovisual input, both with captioned or uncaptioned videos (Bairstow & Lavaur, 2017; Montero Pérez et al., 2013; Peters et al., 2016). With respect to the relationship between semantic relatedness among the items and the number of exposures to them during incidental

learning, our recall data consistently showed that unrelated items benefited more from repeated encounters. Results indicated that recall of the unrelated words increased in a seemingly linear manner with repetitions (see at this regard the linear effect found in the mixed model analysis reported in Footnote 2), while this linear effect was less clear for semantically related words. As a result, we tentatively conclude that the beneficial effects of frequency in L2 vocabulary learning are more plausible to be observed when words are presented in semantically unrelated sets.

In this study, word learning is operationalized by remembering a completely unfamiliar and new label for a previously presented concept (in the images) and L1 word (in the soundtrack). Although unrelated words are not exempt from interference, in the related condition, the existence of semantic competitors is reinforced by the repeated appearance of the elements in the set. In this regard, Finkbeiner & Nicol (2003) suggested that this repetitive and residual activation of concepts and lemmas within the same semantic category could result in a much weaker connection between the new L2 lexical entry and the corresponding concept. With this in mind, we align with the proposal by Wilcox & Medina (2013) suggesting that the lack of effectiveness of semantic clustering could result from an inherent difficulty of the cognitive system to be fed with new words that have already been semantically preorganized, even though the ulterior way of storing accumulated vocabulary could be done in semantic fields. Thus, when designing courses, sessions or textbooks, it is worth considering the potentially detrimental effect of organizing the material in semantic neighborhoods.

An interesting aspect that should be noted is the fact that, unlike Nakata & Suzuki (2018), delayed posttest activities were not included in the current study. In the present study, the negative results found for the semantic clustered set of words in the relatively immediate recall and recognition test phases might be partially explained by the desirable difficulty framework (Bjork, 1999). This theory establishes that the interference caused by semantic clustered conditions makes learning more difficult, but that this difficulty may actually lead to better long-term retention. Further research should be aimed at investigating the effects of semantic clustering and number of encounters in incidental learning situations in the long term to test if undesirable effects of semantic relatedness fade with time.

Previous literature has also suggested that completely unfamiliar or poorly established words are more prone to the interference caused by semantic clustered presentation (Nation, 2000). If this is the case, a possible explanation for results in our study and those from Tinkham (1993, 1997) and Waring (1997) is the use of pseudowords. Using non-existing words ensures the lack of previous knowledge and mimics L2 novice learning situations. However, questions remain about how incidental or semantically related and unrelated words would be in more advanced learners, who are more likely to be familiar with at least one of the related words in the presented set.

There is abundant room for further progress in determining the role of individual differences in vocabulary learning. One relevant issue that emerges directly from our findings is whether a greater reliance on working memory in the learning phase (or enhanced working memory capacity) could moderate the negative effects of presenting new words in semantically related sets. The critical role of this cognitive mechanism has already been shown in incidental vocabulary learning through video (Montero Pérez, 2021), and it has also been demonstrated that working memory has a greater impact when learning occurs intentionally (Bisson et al., 2021). Future research directly comparing intentional and incidental learning scenarios should be oriented at shedding light on the role of working memory in mitigating semantic relatedness effects. Lastly, in future investigations, it might be possible to rotate targets across the two conditions under study (namely, frequency of occurrence

and semantic relatedness). Even though we created two versions of the videos in which the order of appearance of the targets was randomized and participants were also randomly assigned to them, we did control for the effect of extraneous variables since we did not rotate targets across conditions.

In sum, the present study measured the effects of semantic clustering and frequency of appearance on 12 new words presented incidentally and multimodally following a within-participants repeated measures design. Our results showed that words presented in categories which do not fall under the same semantic cluster are better recalled and better recognized than those corresponding to the same semantic neighborhood. Considering the pedagogical implications of these results, we propose that second language teachers and material designers should at least question presenting new vocabulary arranged in groups or lists of words that are semantically related, as the closeness in meaning between the items of each group may burden the learning of the items.

5   References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioral experiment builder. *Behavior research methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Baddeley, A. D. (1997). *Human memory: Theory and practice*. Psychology Press.

Bairstow, D., & Lavaur, J. M. (2017). Sous-titrage, compréhension de films et acquisition de vocabulaire. *Psychologie Française, 62*(3), 249–261. https://doi.org/10.1016/j.psfr.2015.05.001

Barcroft, J. (2015). *Lexical input processing and vocabulary learning*. John Benjamins.

Bisson, M. J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2013). Incidental acquisition of foreign language vocabulary through brief multi-modal exposure. *PLoS ONE, 8*(4). https://doi.org/10.1371/journal.pone.0060912

Bisson, M., Kukona, A., & Lengeris, A. (2021). An ear and eye for language: Mechanisms underlying second language word learning. *Bilingualism: Language and Cognition, 24*(3), 549-568. https://doi.org/10.1017/S1366728920000723

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher and A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (435–459). MIT Press.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*(2), 136–163.

Bürki, A., Elbuy, S., Madec, S., & Vasishth, S. (2020). What did we learn from forty years of research on semantic interference? A bayesian meta-analysis. *Journal of Memory and Language, 114*, 104–125. https://doi.org/10.1016/j.jml.2020.104125.

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics, 31*(5), 693–713. https://doi.org/10.1093/applin/amq031

Semantic clustering on incidental vocabulary learning

Crowder, R. G. (1976). *Principles of Learning and Memory*. Lawrence Erlbaum.

Ellis, R. (1999). *Learning a second language through interaction*. John Benjamins.

Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics, 24*(3), 369–383. https://doi.org/10.1017/S0142716403000195

Folse, K. S. (2004). *Vocabulary myths: Applying second language research to classroom teaching*. University of Michigan Press. DOI: 10.3998/mpub.23925

Grandy, R. E. (1992). Semantic fields, prototypes, and the lexicon. In A. Lehrer and E.F. Kittay (Eds.), *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization* (pp. 105–120). Lawrence Erlbaum.

Guasch M., Ferré P., Fraga I. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods, 48*(4),1358-1369. https://doi.org/10.3758/s13428-015-0684-y

Hashemi, R. H., & Gowdasiaei, F. (2005). An attribute-treatment interaction study: lexical-set versus semantically unrelated vocabulary instruction. *Regional Language Centre Journal*, *36*(3), 341–361.

Haycraft, J. (1993). *An Introduction to English Language Teaching*. Longman.

Higa, M. (1963). Interference effects of intralist word relationships in verbal learning. *Journal of Verbal Learning and Verbal Behavior, 2*(2), 170–75.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11,* 207–223.

Hoshino, Y. (2010). The Categorical Facilitation Effects on L2 Vocabulary Learning in a Classroom Setting. *RELC Journal*, *4*(3)*,* 301–312*.* https://doi.org/10.1177/0033688210380558

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction,* (pp. 258–286). Cambridge University Press.

Hunt, R. R., & Elliot, J. M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General*, *109*(1), 49–74. https://doi.org/10.1037/0096-3445.109.1.49

Hunt, R. R., and Mitchell, D. B. (1982). Independent effects of semantic and nonsemantic distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(1), 81–87. https://doi.org/10.1037/0278-7393.8.1.81

Ishii, T. (2015). Semantic connection or visual connection: Investigating the true source of confusion. *Language Teaching Research*, *19*(6), 712–722. https://doi.org/10.1177/1362168814559799

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods 42*(3), 627-633. https://doi.org/10.3758/BRM.42.3.627

Korochkina, M., Bürki, A., & Nickels, L. (2021). Apples and oranges: How does learning context affect novel word learning? *Journal of Memory and Language, 120*, 1–17. https://doi.org/10.1016/j.jml.2021.104246

Lehrer, A. (1974). *Semantic Fields and Lexical Structure*. American Elsevier.

Malone, J. (2018). Incidental vocabulary learning in SLA: effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition, 40*(3), 651–675. https://doi.org/10.1017/S0272263117000341

Marzano, R. J., & Marzano, J. S. (1988). *A Cluster Approach to Elementary Vocabulary Instruction*. International Reading Association.

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, *41*(3), 720–739. https://doi.org/10.1016/j.system.2013.07.013

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: the role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, *42*(4), 749–773. https://doi.org/10.1017/S0272263119000706

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning and Technology*, *18*(1), 118–141. http://dx.doi.org/10125/44357

Nakata, T., & Suzuki, Y. (2018). Effects of massing and spacing on the learning of semantically related and unrelated words. *Studies in Second Language Acquisition*, *41*(2), 287–311. https://doi.org/10.1017/S0272263118000219

Nation, I. S. P. (2000). Learning vocabulary in lexical sets: Dangers and guidelines. *TESOL Journal*, *9*(2), 6–10.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Heinle Cengage Learning.

Papathanasiou, E. (2009). *An investigation of two ways of presenting vocabulary*. ELT Journal, *63*(4), 313–322. https://doi.org/10.1093/elt/ccp014

Pellicer-Sanchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, *22*(1), 31–55.

Pérez-Serrano, M., Nogueroles-López, M., & Duñabeitia, J.A. (2021). Incidental vocabulary learning with subtitles in a new language: Orthographic markedness and number of exposures. *PLoS ONE*, *16*(2), e0246933. https://doi.org/10.1371/journal.pone.0246933

Peters, E., & Webb, S. (2018). Incidental Vocabulary Acquisition through Viewing L2 Television and Factors that Affect Learning. *Studies in second language acquisition*, *40*(3), 551–577. https://doi.org/10.1017/S0272263117000407

Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, *63*, 134–148. https://doi.org/10.1016/j.system.2016.10.002

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, *18*(1), 1–28.

Rodgers, M. P. H. (2018). The images in television programs and the potential for learning unknown words. The relationship between on-screen imagery and vocabulary. *ITL - International Journal of Applied Linguistics*, *169*(1), 191–211. https://doi.org/10.1075/itl.00012.rod

Rott, S. (1999). The Effect of Exposure Frequency on Intermediate Language Learners' Incidental Vocabulary Acquisition Through Reading. *Studies in Second Language Acquisition*, *21*(4), 589–619.

Sánchez-Gutiérrez, C.H., Pérez-Serrano, M., & Robles-García, P. (2019). The effects of word frequency and typographical enhancement on incidental vocabulary learning in reading. *Journal of Spanish Language Teaching*. 6(1), 14–31. https://doi.org/10.1080/23247797.2019.1590000

Schmitt, N. (2007). Current trends in vocabulary learning and teaching. In J. Cummins & C. Davison (Eds.), *The international handbook of English language teaching* (pp. 827–842). Springer.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.

Seal, B. D. (2001). *Vocabulary learning and teaching.* In M. Celce-Murcia (Ed.), *Teaching English as a Second or Foreign Language* (3rd ed.) (pp. 296–311). Heinle and Heinle.

Stoller, F., Grabe, W. (1995). Implications for L2 vocabulary acquisition and instruction from L1 vocabulary research. In T. Huckin, M. Haynes and J. Coady (Eds.), *Second Language Reading and Vocabulary Learning*, (pp. 24–45). Ablex Publishing Corporation.

Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System*, *21*(3), 371–380. https://doi.org/10.1016/0346-251X(93)90027-E

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research*, *13*(2), 138–163. https://doi.org/10.1191/026765897672376469

Uchihara, T., Webb, S., Saito, K., & Trofimovich, P. (2021). The effects of talker variability and frequency of exposure on the acquisition of spoken word knowledge. *Studies in Second Language Acquisition*, *44*(2), 357–380. https://doi.org/10.1017/S0272263121000218

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening. *System*, *41*(3), 609–624. https://doi.org/10.1016/j.system.2013.07.012

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*(2), 130–163. 10125/66776

Waring, R. (1997). The negative effects of learning words in semantic sets: A replication. *System*, *25*(2), 261–274. https://doi.org/10.1016/S0346-251X(97)00013-4

Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.

Webb, S., & Rodgers, M. P. H. (2009). Vocabulary demands of television programs. *Language Learning*, *59*(2), 335–366. **https://doi.org/10.1111/j.1467-9922.2009.00509.x**

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46–65. https://doi.org/10.1093/applin/aml048

Webb, S. (2020a). Incidental Vocabulary Learning. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 225–239). Routledge.

Webb, S. (2020b). Introduction. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 1–12). Routledge.

Wesche, M., & Paribakht, T.S. (1996). Assessing second language vocabulary knowledge: depth versus breadth. *The Canadian Modern Language Review, 53*(1), 13–40. https://doi.org/10.3138/cmlr.53.1.13

Wharton, S., & Race, P. (1999). *500 Tips for TESOL: Teaching English to Speakers of Other Languages*. Kogan Page Limited.

Wilcox, A., and Medina, A. (2013). Effects of semantic and phonological clustering on L2 vocabulary acquisition among novice learners. *System*, *41*(4), 1056–1069 https://doi.org/10.1016/j.system.2013.10.012

6   Figure

Figure 1. Marginal means plots of the recall accuracy (left), similarity of the recalled strings with the target items (middle) and recognition accuracy (right) split by condition. Error bars represent 95% confidence intervals.

Semantic clustering on incidental vocabulary learning
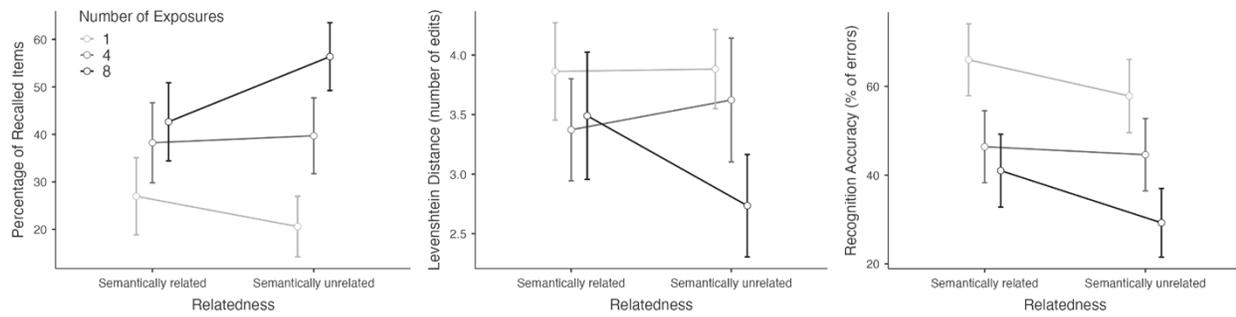
7   Tables

Table 1

Target words and pseudowords used in the experiment, together with the number of appearances of each of them (1, 4, 8), and their corresponding condition with regards to semantic clustering (related, unrelated)

Type of target	Target Word	Pseudoword	Number of Exposures
Semantically unrelated	martillo (hammer)	palvana	8
Semantically unrelated	moto (motorbike)	ina	8
Semantically unrelated	dedo (finger)	pima	4
Semantically unrelated	tenedor (fork)	nacarpa	4
Semantically unrelated	falda (skirt)	marno	1
Semantically unrelated	conejo (rabbit)	brena	1

Semantic clustering on incidental vocabulary learning

Semantically related	uva (grape)	medo	8
Semantically related	manzana (apple)	parcallo	8
Semantically related	pera (pear)	veco	4
Semantically related	naranja (orange)	necador	4
Semantically related	mango (mango)	fesda	1
Semantically related	fresa (strawberry)	cocefa	1

Table 2

Means and standard deviations (in parentheses) of the recall and recognition accuracy data (percentage of errors) per condition

	Semantically related			Semantically unrelated		
	1	4	8	1	4	8
Recall Accuracy (percentage of recalled items)	27 (41.4)	38.2 (42.9)	42.6 (41.9)	20.6 (32.5)	39.7 (40.6)	56.4 (36.3)
Levenshtein Distance (number of edits)	3.86 (2.08)	3.37 (2.18)	3.49 (2.72)	3.88 (1.69)	3.62 (2.65)	2.74 (2.19)
Recognition Accuracy	66.0 (41.2)	46.4 (41.3)	41.0 (41.9)	57.8 (42.1)	44.6 (41.6)	29.2 (39.4)

(percentage of errors)