Cognition 156 (2016) 60–70

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

When pumpkin is closer to onion than to squash: The structure of the second language lexicon

Katy Borodkin [a,*], Yoed N. Kenett [b,1], Miriam Faust [a,b], Nira Mashal [b,c]

[a] Department of Psychology, Bar-Ilan University, Ramat-Gan 52900, Israel
[b] The Leslie and Susan Gonda (Goldschmied) Multidisciplinary Gonda Brain Research Center, Bar-Ilan University, Ramat-Gan 52900, Israel
[c] School of Education, Bar-Ilan University, Ramat-Gan 52900, Israel

ARTICLE INFO

Article history:
Received 15 February 2015
Revised 25 July 2016
Accepted 29 July 2016

Keywords:
Bilingualism
Mental lexicon
Semantic fluency
Small-world network
Modularity

ABSTRACT

The current research investigated the organization of the second language mental lexicon. Twenty-seven English-Hebrew bilingual speakers (who spoke Hebrew as their second language) completed a semantic fluency task in each of their languages, and 24 native Hebrew speakers completed the task in Hebrew. Responses were compared within and across groups, using computational tools. The analyses indicated that the lexical network of the second language displayed greater local connectivity and less modular community structure than the network in the native language, both in the entire sample and in a subsample of bilinguals whose Hebrew vocabulary was matched to that of the native Hebrew speakers. These findings suggest that the lexical network of the second language is not as well-organized as is the network of the first language, even in highly proficient bilinguals. The structural characteristics of the second language lexicon might be affected by factors related to language learning history, including age of acquisition and language use.

1. Introduction

To know a word is to know its spelling and pronunciation, its grammatical class and syntactic constraints, as well as its meaning (Nation, 2001). Word knowledge also refers to usage and associations with other words. A word may have semantic links (e.g., *pumpkin* is likely linked with *zucchini* or with *squash*) as well as associative links with other words with which it tends to co-occur (e.g., *pumpkin* may be linked with *pie*, *orange*, or *Halloween*). Although the structural characteristics of this lexical network have been extensively investigated in monolinguals (e.g., McRae, de Sa, & Seidenberg, 1997; Plaut, 1997), far less is known about the organization of the mental lexicon of bilingual speakers who acquired their languages consecutively. In such individuals, connections are often established first between words in their native language (L1). Second language (L2) words are initially connected only to their L1 translation equivalents; however, they become associated with other L2 words later on, as the L2 vocabulary is acquired, thus giving rise to an autonomous L2 lexical network (Frenck-Mestre & Prince, 1997; Kroll & Stewart, 1994). The current study focuses on the organization of this L2 network in relation to L1 network and explores it with advanced network tools.

Existing research on meaning representation in bilinguals has been largely dedicated to studying cross-linguistic connections through diverse experimental methods, such as cross-language semantic and translation priming, picture naming or Stroop (reviewed in Altarriba & Basnight-Brown, 2009; de Groot, 2011). This line of research is less informative with regard to the connections within the L2 lexical network and the principles governing the organization of this network. Studies addressing the topic more directly have often applied the word association task, in which participants are asked to generate one (or more) associative responses that come to their mind upon presentation of a target word (Kruse, Pankhurst, & Smith, 1987; Söderman, 1993). Responses in this task can be categorized as syntagmatic (words that belong to different lexical classes, such as *pumpkin-orange*), phonological (words that resemble the target word in form but not in meaning, such as *pumpkin-napkin*), and paradigmatic (words that belong to the same lexical class as the target word, such as *pumpkin-squash*). Although there is evidence suggesting that adult L2 speakers, like children in L1 (Ervin, 1961), produce more syntagmatic and phonological responses relative to adult L1 speakers (Meara, 1978; Namei,

* Corresponding author currently at: Department of Communication Disorders, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel.
E-mail addresses: katyborodkin@tauex.tau.ac.il (K. Borodkin), yoedk@sas.upenn.edu (Y.N. Kenett), Miriam.Faust@biu.ac.il (M. Faust), Nira.Mashal@biu.ac.il (N. Mashal).
[1] Present address: Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA.

http://dx.doi.org/10.1016/j.cognition.2016.07.014
0010-0277/

2004; Söderman, 1993), other studies failed to observe differences between L1 and L2 speakers (Kruse et al., 1987; Nissen & Henriksen, 2006). Thus, it is still unclear if and how the lexical structure of L1 and L2 differ. Furthermore, while the word association task is often applied to assess lexical-semantic organization (Kolers, 1963; Van Hell & De Groot, 1998), some researchers have argued that in bilinguals it is also affected by other factors, such as word retrieval difficulties (Antón-Méndez & Gollan, 2010).

Another task that yields inconsistent findings among L1 and L2 speakers is the semantic fluency task. This task is often used in neuropsychological settings and in research to assess language functioning (Ardila, Ostrosky-Solís, & Bernal, 2006). Participants are asked to generate as many different words as possible that belong to a certain category (such as animals and vegetables) in a limited time. Bilinguals performing a semantic fluency task often produce fewer items than monolinguals, both when they are limited to only one of their languages or allowed to use both (Gollan, Montoya, & Werner, 2002; Portocarrero, Burright, & Donovick, 2007; Rosselli & Ardila, 2002; Rosselli et al., 2000; Sandoval, Gollan, Ferreira, & Salmon, 2010). They have been also reported to provide more correct responses in the dominant or more proficient language compared to the nondominant or less proficient language (Sandoval et al., 2010; Taler, Johns, Young, Sheppard, & Jones, 2013). This pattern of results may be explained by between-language interference, which results from the competition between words from both languages that is characteristic of bilingual language production (Gollan et al., 2002; Rosselli et al., 2000; Sandoval et al., 2010). Other researchers, however, did not find significant differences between monolinguals and bilinguals (Bialystok, Craik, & Luk, 2008; Friesen, Luo, Luk, & Bialystok, 2014; Luo, Luk, & Bialystok, 2010) or between the bilinguals' languages (Roberts & Dorze, 1997; Rosselli et al., 2000). These inconsistencies may be accounted in part by differences in vocabulary size (Bialystok et al., 2008). In that study, bilinguals with matched vocabulary scores performed at the same level as monolinguals, and both outperformed bilinguals with lower scores.

Responses on a semantic fluency task can be further analyzed in terms of clustering and switching, and this qualitative analysis has been used as a window into the structure of the bilingual lexicon (Roberts & Dorze, 1997; Rosselli & Ardila, 2002). Clustering refers to the production of sequences of words belonging to the same semantic subcategory, and switching – to the ability to shift to another subcategory (Troyer, 2000; Troyer, Moscovitch, & Winocur, 1997). For example, in the vegetables category, a participant may begin with the squashes family[2] (e.g., *pumpkin*, *squash*, *zucchini*), then switch to the flavorings (e.g., *onion*, *garlic*, *chives*), and so forth until the trial ends. Roberts and Dorze (1997) have demonstrated that clustering measures (i.e., length of clusters and percentage of words in clusters) were greater in French compared to English in bilinguals, indicating richer lexical network in French. The differences, however, were observed only for animal but not food names. The authors speculated that the differences between categories may be related to childhood acquisition patterns of the participants in their study (French-English speakers living in Ottawa, Canada). Similarly, the number of clusters was greater in Spanish than English in older bilinguals living in the US (Rosselli & Ardila, 2002; Salvatierra, Rosselli, Acevedo, & Duara, 2007), suggesting richer network for animal names in L1 (Spanish).

The clustering and switching scoring method (Troyer, 2000; Troyer et al., 1997), however, has faced some criticism. The categorization of responses relies on subjective judgment, which raises potential issues with reliability and validity (Taler et al., 2013). Others question the fundamental assumption of this method that the sequences in semantic fluency responses are indicative of internal lexical-semantic organization (Body & Muskett, 2012). Specifically, Body and Muskett point out the arbitrary nature of classification rules in the existing clustering systems (e.g., Troyer et al. (1997) categorize *camel* as belonging to either *beasts of burden* or *African animals*, but not to *Australian animals*, although camels are native to Australia). Using self-reports, they also show that random factors, such as perceptually salient shared characteristics, rather than semantic organization, determine many of the links between the words in a sequence (e.g., *panda* is followed by *penguin* because they are both black and white).

In the present study, a different approach was taken to explore the structural characteristics of L1 and L2 lexicon, through the use of network science tools. These tools allow for the examination of complex systems (such as the mental lexicon) as web-like structures, or networks, in which *nodes* represent individual entities and *edges* represent links between the entities. The approach has been applied in a variety of domains, including biology, social sciences, and technology (reviewed in Barabási, 2009; Baronchelli, Ferrer-i-Cancho, Pastor-Satorras, Chater, & Christiansen, 2013). In bilingualism research, work within this framework has indicated that the organization of the L2 lexicon is less complex (less dense) than the organization of the L1 lexicon (Wilks & Meara, 2002). However, the authors later admitted that their assumption of many direct connections between words (several dozens) might have been over-simplistic, rendering the conclusions somewhat tentative (Meara, 2009; Wilks, Meara, & Wolter, 2005). The current study further advances this line of research by applying different computational network tools. More specifically, the small-world property and the community structure of L1 and L2 lexical networks are explored.

Networks may be defined in terms of local and global connectivity patterns. In random networks, for instance, local clustering is low (neighboring nodes are sparsely connected to each other), while global distance is short (it takes only a few steps to transverse between distant nodes). Small-world networks, on the other hand, have both high local clustering and short global distance. A network with these characteristics is called a "small-world", because every node in such network is relatively close to almost every other node. Communication transfer in this kind of networks is easy both locally and globally, and thus they are considered optimal (Watts & Strogatz, 1998). More formally, small-world networks are often defined in relation to a random network with the same number of nodes and edges. The comparison is traditionally made based on two parameters, the clustering coefficient (CC) and the average shortest path length (ASPL). The CC refers to the probability that two neighbors of a randomly chosen node will themselves be neighbors, and the ASPL represents the average shortest amount of steps that separate any two pair of random nodes. A small-world network is characterized by having a large CC despite the fact that its ASPL is relatively short and not dramatically different from a random network of comparable size.

At an intermediate level, network organization can be described in terms of community structure (in other words, modularity). A network is considered modular if it has clusters of nodes (communities) that are more densely linked to other nodes within the same community than to nodes outside the community (Newman, 2006). It has been further noted that modular systems tend to be small-world networks, whereas some small-world networks are not necessarily modular (for illustration, see Meunier, Lambiotte, & Bullmore, 2010).

Both the small-world property and modular community structure have been observed in lexical-semantic networks of monolingual adults (Borge-Holthoefer & Arenas, 2010; De Deyne & Storms, 2008; Kenett, Kenett, Ben-Jacob, & Faust, 2011;

[2] Botanically, the squashes family members are fruits, but in common language they are often treated as vegetables. The latter view was adopted for the purposes of this research.