Assessing Second Language (L2) Productive Vocabulary in Health Professionals

Joy Williams, Norman Segalowitz & Tatsiana Leclair
Concordia University & Centre for the Study of Learning and Performance,
Montreal, Quebec

This research is part of a larger investigation into ways of assessing a person's second language (L2) productive vocabulary size. The ultimate goal is to develop a method that can be used in language training of health professionals who need to use their L2 with patients, and in research on language barriers in healthcare communication, where either the health professional or patient is using the L2. The results reported here summarize the study first presented in poster format at the conference and later updated in Williams, Segalowitz & Leclair, in press, *Mental Lexicon*.

We investigated the viability of a novel approach to measuring L2 productive vocabulary size, first proposed by Meara and Almos Alcoy (2010). *Productive* vocabulary can be distinguished from *receptive* vocabulary (Milton, 2009), the former referring to vocabulary a person can produce (speaking, writing) and the latter referring to vocabulary a person can understand (listening, reading). The distinction is sometimes referred to as active versus passive vocabulary. While there is debate about whether productive and receptive vocabulary should be distinguished at all (see Meara, 1997; Melka, 1997), it is generally agreed that there are difficulties developing tasks to assess how many words a person is able to *produce* in a given language. For practical reasons one cannot ask people to write down all the words they know. For theoretical reasons one cannot ask people to identify which words in a list they can produce because that would require word recognition and thus confound receptive with productive vocabulary knowledge. A proposal by Meara and Almos Alcoy (2010) for overcoming these problems is the focus of the research reported here. The approach employs the logic of the capture-recapture method (henceforth CR) used in population ecology (Lindberg & Rexstad, 2002; Petersen, 1896) for estimating how many animals inhabit a given environment or, by analogy, how many words inhabit a person's L2 mental lexicon.

In the CR approach, researchers take samples from the target population at two different times. For example, one would take a representative sample of animals at Time 1, count and tag them for later recognition, and release them back into the environment. At Time 2 the researcher takes a new sample, counts them and notes how many tagged animals have been recaptured. Assuming no change in the population size during the Time 1-Time 2 interval and an even re-distribution of the released animals, then the ratio of the Time 1 sample size (x) to the population size as a whole (p) should be the same as the ratio of the number of recaptured animals (r) at Time 2 to the Time 2 sample size (y). That is, $x/p = r/y$ or $p = xy/r$. By analogy, we created a task to obtain a sample of words at Time 1 and at Time 2, noting how many words from the first sample were repeated or "recaptured" at Time 2. The capture procedure was a free association task in which participants wrote up to 6 words for each of 30 stimulus words selected randomly from across the language. The basic logic is that if people have a small vocabulary, they will run out of words to report and start to repeat themselves while making associations. The more they repeat, the smaller the vocabulary size must be. The goal of the research was to test the viability and validity of the CR approach.

We conducted the CR procedure in participants' first (L1) and second (L2) languages, and collected additional information about their language history and L2 processing skills. The word association data were analyzed in two ways: one was based on the logic of the CR method just described. The other used a different technique to study L2 productive vocabulary, known as the Lex30 method. The Lex30 examines participants' output by counting how many relatively rare words they produce based on frequency norms for the language. The logic of the Lex30 is that if a person's responses contain only the most highly frequent words in the language (say, within the most frequent 2000 words in the language), then that person's vocabulary is relatively low in richness and likely to be relatively small (Meara & Fitzpatrick, 2000; Milton, 2009). We also collected reaction time (RT) data from a living-nonliving visual word judgment task to measure participants' cognitive efficiency in processing word meaning (Segalowitz, 2010).

METHOD

Participants

Participants were 47 English-French bilingual university students (30 females, 17 males; mean age = 23.36, *SD* = 4.07), with varying degrees of self-rated L2 (French) proficiency on a 5-point Likert-type scale. Self-rated L2 proficiency was significantly lower than self-rated L1 proficiency (English).

Materials

Paper-and-pencil word association task. Two sets of 30 different stimulus words were prepared in English and French, drawn from the most frequent 2000 words in each language as indicated in frequency dictionaries (Davies & Gardner, 2010; Lonsdale & Le Bras, 2009). Separate Time 1 and Time 2 booklets each contained one page with 30 English words and sufficient space to write at least 4 word associates, and a similar page with 30 French words.

Living-Nonliving computer-based semantic categorization task. Materials for the reaction time task were 48 words naming living things (e.g., *dog*) and 48 words naming nonliving things (e.g., *bed*) in each language, all different from the stimulus words for the word association task.

Procedure

Participants completed two separate one-hour testing sessions. At Time 1, they did the word association task in their L1 followed by the same task in their L2. For each, they had 15 minutes to write down at least 4 associates to each stimulus word. They then completed the living-nonliving task in the L1 and L2 by pressing the appropriate reaction time panel to indicate whether a word shown on the computer screen named a living or a non-living thing. They also did the first half of a language background questionnaire. A few days later, at Time 2, the participants again did the L1 and L2 word association tasks, with different sets of 30 stimulus words and completed the language background questionnaire.

MAIN RESULTS

We calculated CR scores in each language as the number of different word associates generated at Time 1 (x) times the number of different word associates generated at Time 2 (y) divided by the number of Time 1 words repeated in the Time 2 responses (r) ($p = xy/r$). Lex30 scores in each language at Time 1 were obtained by counting the number of word associates (lemmas) not found in the 2000 most frequent words in the language. All L2 measures were residualized against L1 measures to correct for general individual differences in test performance that would be reflected in L1 performance.

The main results were the following:

(1) L1 (English) CR and Lex30 scores correlated significantly ($r_s = .66$, $p < .001$), as did the residualized L2 (French) CR and Lex30 scores (also $r_s = .66$, $p < .001$), indicating convergent validity for the CR method.
(2) CR measures in L1 were significantly greater than in L2 (unresidualized) (889.97 and 606.67 respectively; Wilcoxon signed-rank test, $p < .001$), supporting construct validity for the CR method. The Lex30 measure failed to discriminate between the L1 and L2 vocabulary size (Wilcoxon signed-rank test, $p = .30$).
(3) Residualized L2 (French) CR scores correlated significantly with residualized L2 RTs from the Living-Nonliving task ($r_s = -.44$, $p < .01$), supporting construct validity for the CR method (larger vocabulary associated with faster processing).

CONCLUSIONS

The CR estimate is a valid index of L2 productive vocabulary size. The word association test is appropriate for use within this paradigm. The CR index is not, however, a reflection of the *absolute* L1 and L2 vocabulary size.

ACKNOWLEDGEMENT: This research was funded by Health Canada and administered by the McGill Training and Retention of Health Professionals Project.

REFERENCES

Fitzpatrick, T., & Meara, P. (2004). Exploring the validity of a test of productive vocabulary. *Vigo International Journal of Applied Linguistics*, *1*, 55–73.

Lindberg, M., & Rexstad, E. (2002). Capture-recapture sampling designs. In A. H. El-Shaarawi & W. W. Piegorsch. *Encyclopedia of Environmetrics, Volume* 1. (pp. 251- 262). Chichister: John Wiley & Sons.

Meara, P. (1997). Towards a new approach to modelling vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition, and Pedagogy* (pp. 109-121). Cambridge: Cambridge University Press.

Meara, P., & Fitzpatrick, T. (2000) Lex 30: An improved method for assessing productive vocabulary in an L2. *System*, *28*, 19–30.

Meara, P. M., & Olmos Alcoy, J. C. (2010). Words as species: An alternative approach to estimating productive vocabulary size. *Reading in a Foreign Language*, *22*, 222–236.

Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt & M. McCarthy, (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 84-102). Cambridge, MA: Cambridge University Press.

Petersen, C. G. J. (1896). The yearly immigration of young plaice into the Linsfjord from the German Sea. *Report of the Danish Biological Station*, *6*, 5–84.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol, UK: Multilingual Matters.