Which words do English non-native speakers know? New supernational levels based on yes/no decision

Marc Brysbaert[1], Emmanuel Keuleers[2], Paweł Mandera[1]

[1] Department of Experimental Psychology, Ghent University

[2] Department of Cognitive Science and Artificial Intelligence, Tilburg University

Keywords: Word prevalence, word frequency, word knowledge, megastudy

To be published in *Second Language Research*

Address: Marc Brysbaert
Department of Experimental Psychology
Henri Dunantlaan 2
B-9000 Gent
Belgium
Tel. +32 9 264 94 25
Fax. +32 9 264 64 96
E-mal: marc.brysbaert@ugent.be

Abstract

To have more information about the English words known by L2 speakers, we ran a large-scale crowdsourcing vocabulary test, which yielded 17 million useful responses. It provided us with a list of 445 words known to nearly all participants. The list was compared to various existing lists of words advised to include in the first stages of English L2 teaching. The data also provided us with a ranking of 61 thousand words in terms of degree and speed of word recognition in English L2 speakers, which correlated r = .85 with a similar ranking based on native English speakers. The L2 speakers in our study were relatively better at academic words (which are often cognates in their mother tongue) and words related to experiences English L2 students are likely to have. They were worse at words related to childhood and family life. Finally, a new list of 20 levels of 1000 word families is presented, which will be of use to English L2 teachers, as the levels represent the order in which English vocabulary seems to be acquired by L2 learners across the world.

Keywords: English L2 word knowledge, language acquisition, vocabulary

How many words do we need for language understanding?

A typical 20-year old native speaker knows some 42 thousand base words (lemmas) derived from 11,100 word families (Brysbaert, Stevens, Mandera, & Keuleers, 2016). This is a hefty challenge for someone wanting to learn a new language (a so-called L2 – second language – compared to L1 –first language or mother tongue). Cobb (2007, 2016) argued that many L2 learners will only acquire the 2,500-3,000 most frequent word families of the language. A word family is defined as a word with its inflections and transparent derivations (e.g., *play, playing, plays, player, …;* see Bauer & Nation, 1993, for more information). The 2,500 – 3,000 most frequent word families occur regularly enough to be picked up during typical L2 conversations. The other word families are too infrequent to be learned on the basis of exposure alone. According to Cobb, these word families must be taught explicitly if they are to be acquired.

Evidence in line with Cobb's argument was published by Webb and Chang (2012). They followed 166 Taiwanese learners of English as a foreign language for five years from the last years of high school into university. In the final year of the study only 47% of the participants mastered the 1,000 most common word families, and 16% mastered the next level of 2,000 often-used word families.

Nation and Waring (1997) estimated that a person with knowledge of the 2,000 most frequent word families understands most of the words used in familiar social interactions, but would miss one word out of five in written texts. This makes it nearly impossible to understand the text. For written text understanding, Nation and Waring ventured that a minimum of 5,000 word families is necessary, which cover some 95% of the words (i.e., one word in 20 unknown). Laufer and Ravenhorst-Kalovski (2010) later proposed 8,000 word families as a better threshold for text understanding (98% coverage; see also Nation, 2006).

The observation that not all words must be known for successful language understanding illustrates that some words are more useful than others. An obvious criterion in this respect is word frequency: the number of times a word (or word family) is used in the language. It is more helpful to know the English verb *to go* than the verb *to beseech*, because you are much more likely to encounter the former in the English language.[1]

Lists of "important" words to learn

Several lists have been compiled of the first words families to be learned (taught) in English.[2] West (1953), for instance, published the General Service List (GSL), consisting of the 2,284 most important word families, which were made available by Bauman (see http://jbauman.com/gsl.html). New lists to replace the GSL were created 60 years after the initial publication independently by Browne, Culligan, and Phillips (see Browne, 2014; with 2,368 word families) and Brezina and Gablasova (2015; with 2,494 word families).

Other lists of basic words focused on non-native English students going to English-speaking universities. Coxhead (2000), for instance, published an academic word list of 570 word families accounting for 10% of the total words (tokens) in academic texts but only 1.4% of the total words in a fiction corpus of the same size. Gardner and Davies (2014) published an updated academic vocabulary list of 1991 word families.

More ambitious attempts involved the classification of "all" worthwhile word families in terms of usefulness. Arguably the best known is Nation's (2006) division of English words into 14 levels of 1,000 word families with decreasing word frequencies. In the latest version of Nation's list, a distinction was made between 26 levels of 1,000 families each (available at https://www.victoria.ac.nz/lals/about/staff/paul-nation). Another example of word classification for L2 teaching is the JACET list of 8 levels with 1,000 word families each, compiled for Japanese learners of English (Uemura & Ishikawa, 2004).

Cambridge University Press published the English Profile Wordlists for English L2 learners at various stages within the Common European Framework of Reference (Capel, 2010). The

[1] Notice that word frequency is often disregarded in L2 teaching. This is the case, for instance, when students must learn a list of irregular verbs in which *to go* and *to beseech* are given the same weight.

[2] Such efforts are not limited to the English language. Similar initiatives were made for other languages, such as "le français fundamental" in the 1950s for the French language (Gineste & Lagrave, 1961).

Common European Framework of Reference (CEFR) divides language proficiency into six levels:

- A1: Can understand and use a few familiar everyday expressions.
- A2: Can communicate in simple and routine tasks.
- B1: Can deal with situations that are familiar and of personal interest.
- B2: Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers possible without strain for either party.
- C1: Can express ideas fluently and spontaneously in a multitude of contexts without much obvious searching for expressions.
- C2: Close to native language use; can summarize information from different spoken and written sources, can reconstruct arguments and accounts in a coherent presentation.

The selection of words for the English Profile Wordlists was based on both word frequency and the words used in English tests at the different levels. A word like *brother* is supposed to be known at stage A1, whereas the word *bundle* is only expected to be known at level C2. In total, the English Profile Wordlists provide a classification for 15 thousand meanings/senses[3] of words and expressions. Unlike Nation's list and the JACET list, words and expressions can have more than one meaning/sense. For instance, the meanings/senses "go with" and "happen together" of the verb *accompany* are supposed to be known at level B1, whereas the meaning/sense "to provide a musical accompaniment for a performer" is only expected to be known at level C2.

A similar initiative was introduced by Pearson Education (Benigno & de Jong, 2019), whose *Global Scale of English* provides information for 38 thousand meanings of words and expressions. According to this list, the word *brother* must be known at A1 when it refers to a family member and at C1 when it refers to fellow believers. In contrast, all meanings of the verb *to accompany* are supposed to be acquired at B2.

[3] The term meaning is generally used for unrelated interpretations (e.g., bank as financial building vs. bank as the side of a river), whereas sense refers to related interpretations (e.g., bank as a building vs. bank as a financial institute). If a word has different meanings the interpretations cannot be derived from each other, whereas this is possible for a word with several senses. As is often true for natural concepts, the border between meanings and senses is fuzzy.

A final list linking English words to CEFR levels is the EFLLex list compiled by a group of European universities (Dürlich & François, 2018). It contains information about 15,280 words observed in the materials developed for teaching at the different CEFR proficiency levels. According to this list, the word *brother* (no meaning distinction made) must be known at level A1. *Accompany* is unlikely to be known before level C1.

Top-down selection vs. bottom-up assessment

All lists just described were compiled top-down. Groups of researchers sat together and decided "to the best of their knowledge" which words should be known when. As indicated, the main criterion relied upon was word frequency, with high frequency words given prominence over low frequency words. Because there are different frequency lists and because there is freedom in other criteria to use, the lists differ to some extent in their classification of words, as we have seen above.

Word frequency is an important variable to determine word usefulness, but has two limitations as the sole indicator of word usefulness (He & Godfroid, 2019; Laufer & Nation, 2012). The first is that word frequency forms a very asymmetric curve. There are few high-frequency words (2,000 word families is a reasonable ballpark figure) and the vast bulk of word families are part of a long tail of low-frequency words. As a result, categorization works well for the first two or maybe three levels of 1,000 words, but rapidly becomes arbitrary afterwards, depending on the corpus used to determine word frequencies.

A second limitation is that some low-frequency words are likely to be known by the vast majority of L2 learners, whereas others are not. Take the words *noun*, *verb*, and *vocabulary*. They all have low frequencies (about 1-2 occurrences per million words), but everyone learning English in class is likely to know them. The same is true for words describing other important elements in the world of English L2 learners. In contrast, some words known by most native speakers are unlikely to be known by L2 speakers, because they do not belong to the typical world of an L2 speaker. This is the case for words related to the world of children, such as *fib*, *thimble*, or *poppycock*. It may be difficult for researchers to estimate the optimal level of these words, given that their intuitions differ from those of L2 learners.

So, an interesting alternative is to examine word usefulness bottom-up, to assess which words are known by bilinguals, independent of word frequency or researcher intuition. This can be done by analyzing the data from large-scale vocabulary studies. Keuleers, Stevens, Mandera, and Brysbaert (2015) defined knowledge of a word in a particular population as **word prevalence** and they published two articles on English word prevalence norms for native speakers (Brysbaert, Mandera, McCormick, & Keuleers, 2019b; Mandera, Keuleers, & Brysbaert, 2020).

In the present article we describe word prevalence norms for English L2 speakers. Which are the best known English words? And how well are various words known to L2 speakers? The English test grew out of a similar initiative in Dutch. The aim was to compile a large list of words based on dictionaries and corpus analysis, and to see how many people know each word. The words were presented via an easily accessible, short internet test to the general public. Each participant saw a small sample of words and were asked to indicate which words they knew. To discourage participants from saying ‘yes’ to words they did not know, a subsample of the stimuli were made-up, non-existing words. If participants indicated they knew these ‘words’, their vocabulary scores were corrected for false alarms. Although word recognition is measured at a shallow level (do you recognize this word?), many studies have indicated that the test scores correlate well with those of more demanding language tests that probe language knowledge in depth (Ferré & Brysbaert, 2017; Harrington, & Carey, 2009; Lemhöfer & Broersma, 2012; Meara, & Buxton, 1987; Zhang, Liu, & Ai, 2019).

Method

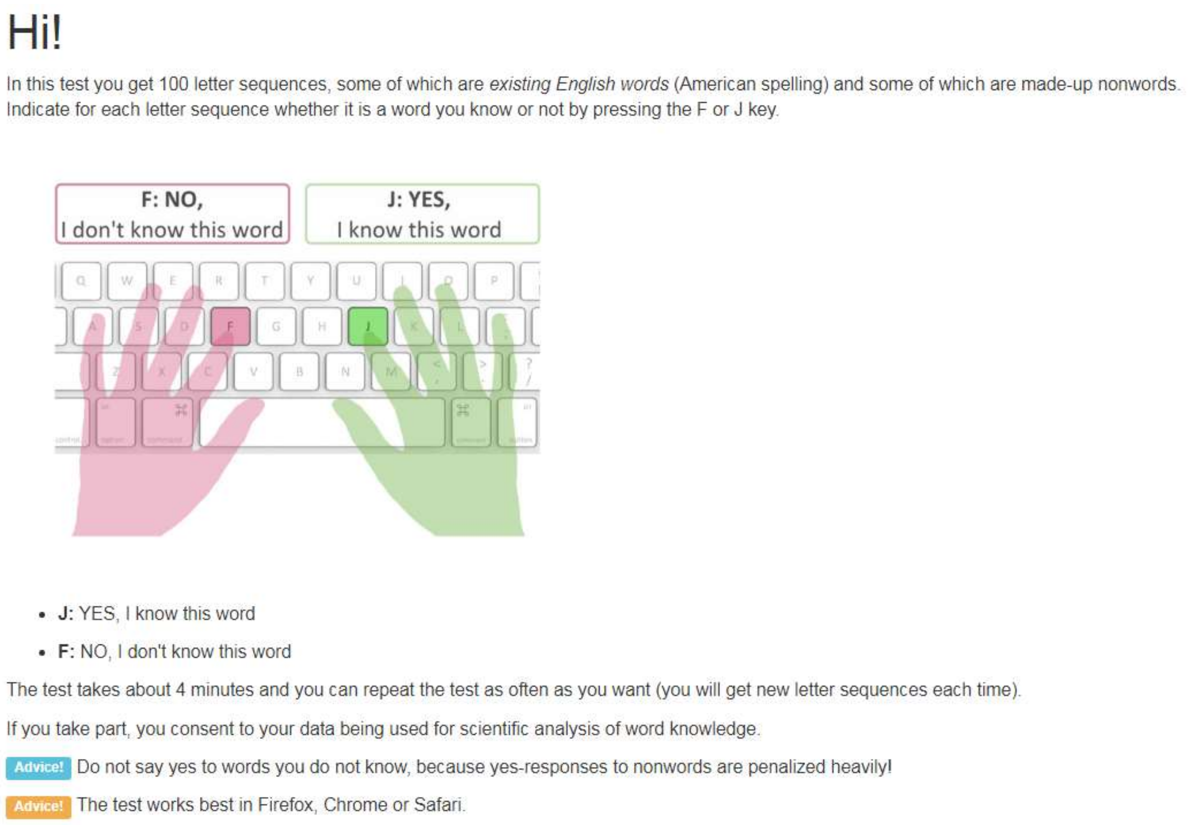
Stimulus materials

The list of stimulus materials used consists of 61,851 English words. The vast majority of them are lemmas (the base, dictionary forms of the words). A few are irregular word inflections, such as *lice* or *been*. Inflected verb forms are included as separate entries if they are often used as nouns or adjectives. Examples are *finished*, *tuning* or *undeveloped*. In addition to the word list, we used an extensive list of non-words. These are sequences of letters that could form an acceptable English word but that do not exist. Examples are *sulched*, *pawler*, or *miptly* (see Keuleers & Brysbaert, 2010, for more information on how the non-words were made). The non-words were included to make sure that participants were not tempted to say “yes” to words they did not know.

Procedure

The test (which is still running) was administered via the internet (http://vocabulary.ugent.be/). For each test event taken, a new random sample of 70 words and 30 non-words was selected (i.e., participants could do the test multiple times, always getting a different sample of stimulus materials).

The test started with the screen shown in Figure 1. [4]



Hi!

In this test you get 100 letter sequences, some of which are *existing English words* (American spelling) and some of which are made-up nonwords. Indicate for each letter sequence whether it is a word you know or not by pressing the F or J key.

- **J:** YES, I know this word
- **F:** NO, I don't know this word

The test takes about 4 minutes and you can repeat the test as often as you want (you will get new letter sequences each time).

If you take part, you consent to your data being used for scientific analysis of word knowledge.

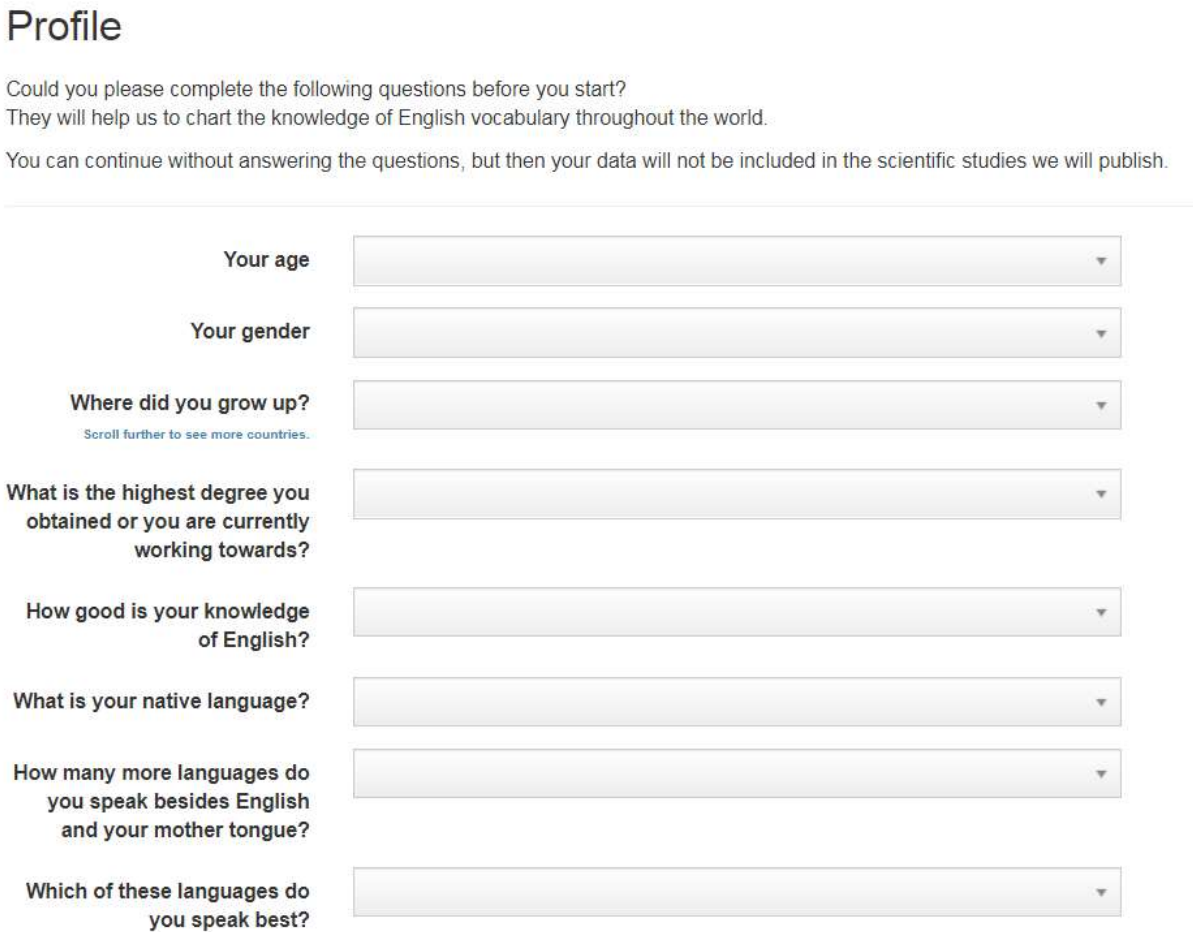
Advice! Do not say yes to words you do not know, because yes-responses to nonwords are penalized heavily!

Advice! The test works best in Firefox, Chrome or Safari.

Figure 1: Opening screen of the vocabulary test, explaining the nature of the test to the participants.

[4] The test automatically adapted to the screen used (e.g., desktop computer or cell phone). As a result, the lay-out changed (and the way of collecting data, e.g. via a touch screen), but the information provided was always the same.

Next, we asked some elementary personal information (Figure 2). Participants could opt not to respond to these questions and were informed that their data would not be used for analysis if they did so. Particularly important for the present purpose are the questions about the native language and how participants estimated their proficiency.



Profile

Could you please complete the following questions before you start?
They will help us to chart the knowledge of English vocabulary throughout the world.

You can continue without answering the questions, but then your data will not be included in the scientific studies we will publish.

Your age

Your gender

Where did you grow up?
Scroll further to see more countries.

What is the highest degree you obtained or you are currently working towards?

How good is your knowledge of English?

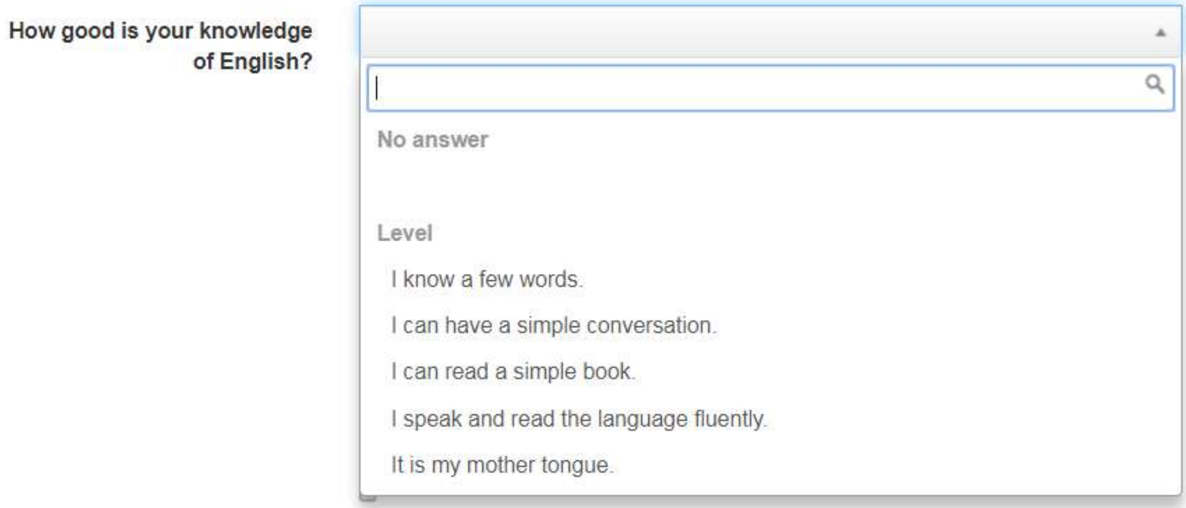
What is your native language?

How many more languages do you speak besides English and your mother tongue?

Which of these languages do you speak best?

Figure 2: Person-related questions asked to the participants.

With respect to language proficiency, participants could choose between six levels, inspired on CEFR (Figure 3).



How good is your knowledge of English?

No answer

Level

I know a few words.

I can have a simple conversation.

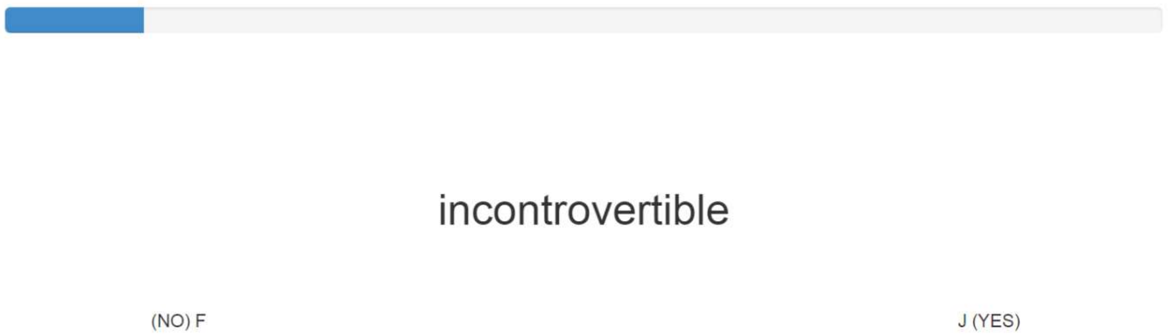
I can read a simple book.

I speak and read the language fluently.

It is my mother tongue.

Figure 3: English proficiency levels the participants could choose from

On each trial, participants saw a string of letters and had to indicate whether it formed an English word they knew (Figure 4).



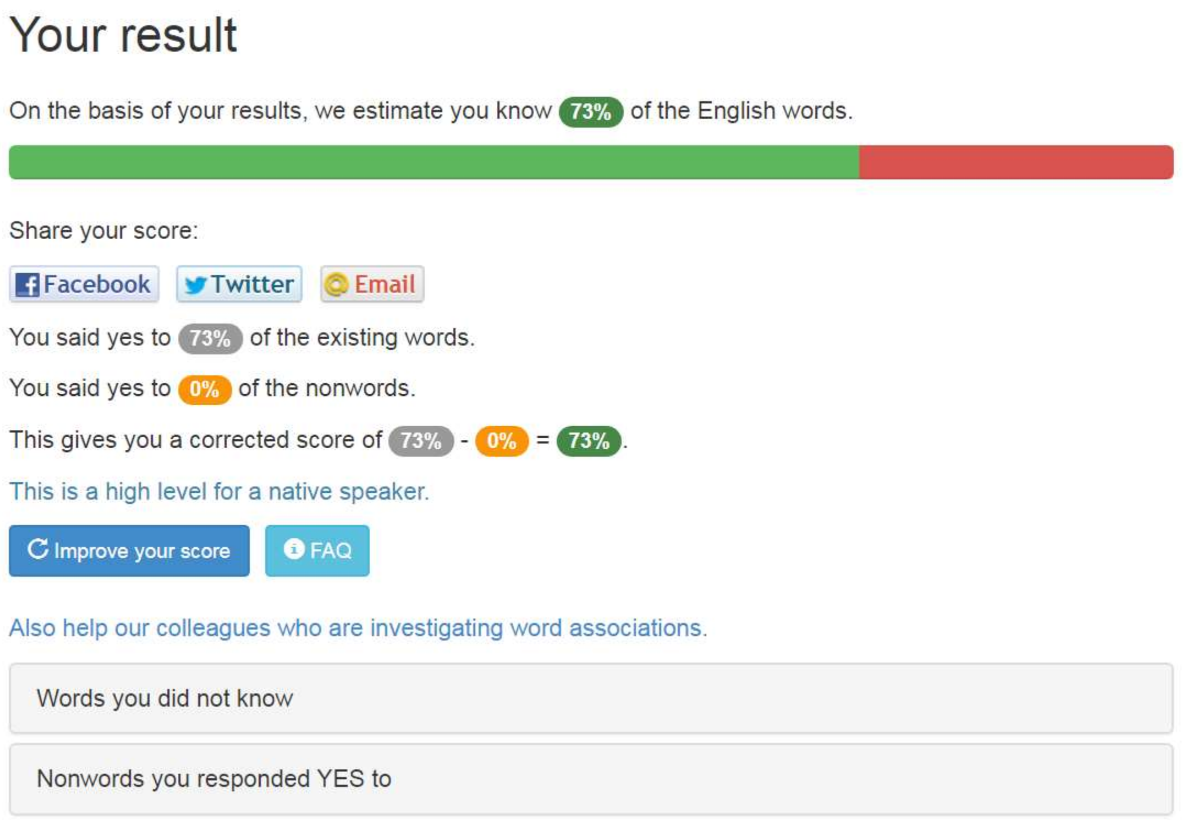
incontrovertible

(NO) F

J (YES)

Figure 4: Screenshot of a trial. Participants saw a sequence of letters and had to indicate whether it was an English word they knew. If so, they pressed a key with their right hand; if not, they pressed a key with their left hand. A progress bar at the top indicated how much of the session had been done already. A session of 100 trials typically took 3-4 minutes.

At the end, participants received feedback about their performance (Figure 5), which was a big motivator to do the test and share it with others. It gave the percentage of yes-responses to words (out of 70), the percentage of yes-responses to non-words (out of 30), and the corrected score (words minus non-words). By including non-words (and telling participants they would be penalized if they wrongly selected them as known English words), we discouraged participants from selecting words they did not know.



Your result

On the basis of your results, we estimate you know 73% of the English words.

Share your score:

You said yes to 73% of the existing words.

You said yes to 0% of the nonwords.

This gives you a corrected score of 73% - 0% = 73%.

This is a high level for a native speaker.

Also help our colleagues who are investigating word associations.

Words you did not know

Nonwords you responded YES to

Figure 5: Feedback provided to the participants at the end of the test.

After the test, participants could look up the words they did not know (with a link to a dictionary) and check the non-words they erroneously thought they knew. If they wanted to, they could take the test again (with different stimuli) and share their results on social media.

Publicity for the test was limited to sending a link to a few popular news sites and influencers, some of whom shared the link. The only reward was that participants got an estimate of their mastery of English words, which they could compare to that of friends. In a few weeks, however, word of mouth made the test take off to numbers not imagined at the offset. Even in 2020, six years after the launch, the test is completed by some 500 users each day. Also the diversity of participants turned out to be much larger than the people we can be reach in typical university-based studies (although still not representative for the total population).

Results

When we set up the vocabulary study, we expected to get useful information about word knowledge (i.e., how many participants pressed yes to each word). We also measured the time the participants took to make their response for exploratory purposes. To our pleasant surprise, these reaction times turned out to correlate .7 with reaction times collected under laboratory-controlled lexical decision tasks (Brysbaert, Keuleers, & Mandera, 2019b; Mandera, Keuleers, & Brysbaert, 2020). So, they can be used as an indication of how easy it is to recognize a word.

We followed the same pipeline of cleaning procedures as in Brysbaert et al. (2019b) and Mandera et al. (2020). They are:

1) We only took into account the word data of the people who answered the person related questions.
2) We only used the first 3 sessions from each IP-address, to make sure that no individual had an undue influence (some participants did hundreds of sessions).
3) For reaction times we deleted for the first 9 trials of each session, which were considered training trials.
4) Trials with reaction times longer than 8000 ms were deleted, so that no dictionary consultation could take place.
5) Reaction time outliers were further filtered out based on an adjusted boxplot method for positively skewed distributions calculated separately for the words in each individual session.
6) In addition to raw reaction times, we also calculated standardized reaction times. These are the reaction times expressed as z-scores calculated per session. Such z-scores result in more reliable estimates of word processing times, because speed differences between participants (and sessions of individual participants) are partialed out. The percentage of

variance explained by word variables typically is some 10% more for standardized reaction times than for raw reaction times.

7) Only data from users who indicated that English was not their native language were retained. This includes any answer other than “English” to the question about the native language and other than “It is my mother tongue” to the question regarding the knowledge of English.

For the present analyses, we additionally omitted sessions with more than 2 yes-responses to nonwords, so that the data are relatively free from guessing on the basis of “word likeliness” (a maximum of 7% guessing).

The pruning reduced the initial sample of 463 thousand sessions from 385 thousand non-native English speakers to 286.5 thousand useful sessions (almost 17 million answers to words).

More than 150 different mother tongues were indicated by the participants. Of these, 42 had more than 1000 participants in the final dataset. The largest groups were Spanish (33.3 thousand sessions), Hungarian (32.3 thousand sessions), German (19.8 thousand sessions), Polish (18.4 thousand sessions), Dutch (17.4 thousand sessions), and Chinese (13.4 thousand sessions).[5]

Mean age of the participants was 30 years (SD = 11.4). Of them, 2.4% of the participants indicated that the highest degree obtained in school was primary school; for 21.5% it was high school; for 37.8% bachelor; for 29.9% master; and for 8.4% PhD.

Participant differences in L2 word knowledge

Table 1 shows the number of sessions in the six proficiency levels and the average performance scores obtained by the participants. It clearly illustrates the growth of vocabulary as participants become more proficient in a language. One percent represents

[5] Although it is tempting to break down the discussion for different language groups, it is important to know that the numbers of observations per word rapidly become too small to have stable results. Indeed, the strength of the overall analysis is that it is based on 200-300 observations per word.

618 words. So, the mean number of lemmas known by the native group is 74 * 618 = 46 thousand.[6] The number of words known by participants who indicated they knew a few words, was surprisingly high: 35 * 618 = 21 thousand words. We will return to this observation in the discussion section.

Level	Nsessions	Score
"I know a few words"	7,104	0.35 (SD=0.20)
"I can have a simple conversation"	21,676	0.37 (SD=0.17)
"I can read a simple book"	58,269	0.45 (SD=0.17)
"I speak and read the language fluently"	199,398	0.62 (SD=0.15)
"It is my mother tongue"[7]	431,924	0.74 (SD=0.11)

Table 1: Number of sessions and performance for the 6 proficiency levels. Data for the native speakers come from Brysbaert et al. (2016).

The best known words

More interesting than the analysis over participants for the current article is the analysis over items. How well are various words known? Each word was judged on average 274 times, giving us fairly stable estimates.

For a start, we can look at the words selected by everyone. These are the words known to all English L2 speakers (at least those who participated in our study). A total of 114 words were recognized by all participants. We loosely divided them into eight categories that may be salient for English L2 speakers[8]: Persons involved in conversations, actions performed, things, attributes of things and people, indications of place and time, conversation topics,

[6] The number increased from 42K for 20-year olds to 48K for 60-year olds (Brysbaert et al., 2016).

[7] The selection criterion was "It is my mother tongue" indicated by the participant in the question about the level of English, and "English" specified as a native language in the respective field.

[8] Readers who disagree with our classification, are invited to make other, more theory-based categorizations.

function words, and language-class-related words. Table 2 shows the distribution of the 114 words over the categories.

Table 2: 114 words known by all the English L2 speakers (ordered according to speed of responding, which is an indication of how easy the word is)

Persons	Group, me, player, actor, you, men, woman, who, children, kids, someone, manager, somebody, uncle, cleaner, us, secretary
Actions	Help, smile, start, phone, eat, walk, find, think, read, move, save, do, check, hate, telephone, promise, finish, believe, crying, snowboarding
Things and animals	Coffee, water, music, radio, problem, sun, sugar, toy, rule, sport, bath, milk, party, horse, level, foot, song, shirt, hotel, service, document, present, oxygen, motivation, address, biology, darkness, corner, optimism, chocolate, magazine, technology, popularity, experiment, intelligence
Attributes	Best, full, amazing, sexy, happy, yellow, global, big, born, right, positive, five, back, two, broken, this, online, finished, expensive, officially, said, historic, surviving, seventeen
Place and time	Room, city, soon, day, sky, hospital, tomorrow, airport, midnight
Topics	History
Function words	Hello, between, often, inside, how
Language-class-related	Subject, verb, vocabulary

Interestingly, quite a few of the words are not expected on the basis of word frequency. The knowledge of them is more likely to come from English classes, traveling, internet interactions, watching television, and playing games. Some words are also cognates in several languages (*oxygen, motivation, intelligence*).

Table 3 shows the presence of the words in various lists we discussed.[9] A few of the words are not included in the lists, because they were considered as inflected verb forms rather than as adjectives or nouns (*crying, broken, finished*). Other words are underestimated because they have a low frequency in written texts (*verb, midnight, secretary*). Finally, all but Nation's lists seem to underestimate the knowledge of derived forms (*sexy, darkness*). A reassuring observation is that the two new General Service Lists include more of the words than the original list (see also Dang, Webb, & Coxhead, 2020).

Table 3: Inclusion of the words known to all L2 participants in various lists of useful words (GSL = General Service List, NGSL_Browne = New General Service List Browne et al; NGSL_B&G = New General Service list Brezina & Gablasova; AWL = Academic Word List; Nation = level in Nation's Range program; JACET = level in the JACET list; Global Scale = CEFR level in the Global Scale of English; Profile = CEFR level in the English Profile Wordlists). The list is ordered according to the speed with which the participants selected the words, as this is an indication of how easy the words are. List 33 of Nation includes transparent compound words.

Word	**GSL**	**NGSL_Browne**	**NGSL_B&G**	**AWL**	**Nation**	**JACET**	**Global scale**	**Profile**
help	x	x	x		1	1	A1	<A1
best	x				1	1	A1	A1
smile	x	x	x		1	1	B1	A2
full	x	x	x		1	1	A2	A1
coffee	x	x	x		1	2	A1	A1
water	x	x	x		1	1	A1	<A1
room	x	x	x		1	1	A1	<A1
music	x	x	x		1	1	A1	A1
amazing					1	3	A2	A2+
history	x	x	x		1	1	A2	A2
sexy					1	7	B2	B2
happy	x	x	x		1	1	A1	<A1

[9] EFLLex is not included in the table, because it does not include specific levels. It provides information about word frequency and presence in various sources.

city	x	x	x		1	1	A1	<A1
yellow	x	x	x		1	2	A1	A1
soon	x	x	x		1	1	A2	A2
start	x	x	x		1	1	A1	<A1
phone		x	x		1	1	A1	<A1
global		x	x		3	2	B2	A2+
radio	x	x	x		1	1	A1	<A1
group	x	x	x		1	1	A1	A1
problem	x	x	x		1	1	A1	A1
day	x	x	x		1	1	A1	<A1
sun	x	x	x		1	1	A1	B1
me					1	1	A1	<A1
player		x	x		1	1	A1	A2
sugar	x	x	x		2	2	A1	A1
eat	x	x	x		1	1	A1	<A1
walk	x	x	x		1	1	A1	<A1
sky	x	x	x		1	1	A2	A1
big	x	x	x		1	1	A1	<A1
hospital	x	x	x		1	1	A1	A2
born					1			A2
toy	x	x			2	3	A2	A2+
right	x	x	x		1	1	A1	<A1
actor	x	x	x		1	2	A2	A2
rule	x	x	x		1	1	B1	A1
positive		x	x	x	2	2	B1	B1
find	x	x	x		1	1	A1	A2
think	x	x	x		1	1	A1	<A1
you	x	x	x		1	1	A1	<A1
sport	x	x	x		1	1	A1	<A1
read	x	x	x		1	1	A1	<A1
hello	x	x			1	2	A1	<A1
bath	x	x			1	2	A1	A1
five			x		1		A1	N/A
milk	x	x	x		1	2	A1	A1
move	x	x	x		1	1	A2	A2
party	x	x	x		1	1	A1	A2
men					1			
horse	x	x	x		1	1	A1	<A1
woman	x	x	x		1	1	A1	<A1
level	x	x	x		1	1	A2	B1
save	x	x	x		1	1	A2	A2
tomorrow	x	x	x		1	1	A1	<A1
foot	x	x	x		1	1	A1	<A1
back	x	x	x		1	1	A1	<A1
two			x		1		A1	N/A
airport			x		33	2	A2	<A1
do	x	x	x		1	1	A1	<A1
who	x	x	x		1	1	A1	A2

song	x	x	x		1	1	A2	<A1
shirt	x	x	x		1	2	A1	A1
children					1			<A1
hotel	x	x	x		2	1	A1	<A1
subject	x	x	x		1	1	A1	A2
broken					1	2	A2	B1
service	x	x	x		1	1	B1	A2
kids					1			
document		x	x	x	3	3	A2	B1
check	x	x	x		1	1	A2	A2
present	x	x	x		1	1	A2	A2
this	x	x	x		1	1	A1	<A1
between	x	x	x		1	1	A1	A1
often	x	x	x		1	1	A1	A2
online		x	x		33	5	A2	A2
hate	x	x	x		1	2	A2	A2
telephone	x	x	x		1	1	A2	A1
finished					1	6		B1+
promise	x	x	x		1	1	B1	B1
oxygen					4	3	B2	B1
verb	x	x			5	5	A2	A2+
someone	x	x	x		1	1	A2	<A1
motivation		x			3	4	B2	B2+
inside	x	x	x		1	1	A1	A1
expensive	x	x	x		1	2	A1	<A1
address	x	x	x		1	1	A1	<A1
finish	x	x	x		1	1	A1	<A1
biology					4	5	A2	B2
how	x	x	x		1	1	A1	<A1
crying					1			
darkness		x			1	2	B2	B2
believe	x	x	x		1	1	A2	A2
corner	x	x	x		1	1	A2	A2
manager		x	x		1	1	A2	A2
somebody	x	x	x		1	2	A2	<A1
optimism					5	6	C2	B2+
midnight					33	3	A2	B1
uncle	x	x			1	2	A2	A2+
chocolate		x	x		2	3	A1	A1
magazine		x	x		2	2	A2	A2+
technology		x	x	x	2	1	B1	A2
popularity					2	5	B2	B2
officially					2	4	C1	B1+
cleaner					1	5	A2	B1+
said					1			
historic		x	x		1	3	B1	B1+
vocabulary					5	5	A2	A2+
us					1	1	A1	<A1

experiment	x	x	x		3	2	B1	B1
surviving					2			
intelligence		x	x	x	3	2	B2	B1+
seventeen					1		A1	N/A
secretary	x	x			3	2	A2	B1
snowboarding					33		A2	B2+

There are 331 words with one "no" response. Of course, the distinction between zero and one "no" response has an element of arbitrariness (slip of attention, number of participants responding to the word, proficiency level of the people seeing the word, response bias, etc.). Still, the criterion is valid, because as a group words with one error are slightly more difficult than words with no errors.

In the supplementary materials, an Excel file is available with all 445 words that elicited zero or one no-response, together with the information from the lists in Table 3. This file is also available on https://osf.io/gakre/. Table 4 shows how many words are present in the various lists. Of the General Service Lists, Browne et al.'s does best, followed by Brezina and Gablasova's, and the original list. Only 16 words are in the Academic Word List. Nation's list has the highest coverage and the best correspondence. The words most misjudged are: *cellphone, subway, vampire, shampoo, password, disconnection, optimism, paradise, microphone, basketball, vocabulary*, and *verb*. Three words are not in Nation's list: *T-shirt, smartphone*, and *facebook*. Next to the words misjudged in Nation's lists, JACET seriously misjudges knowledge of derived words: *unsafe, unlucky, speedy, incorrect,* and *sexy*.

Specifically related to CEFR, both the Global Scale of English and the English Profile Wordlists have only one third of the generally known English words in the A1 level. Almost one third are outside the A1+A2 level or are missing in the lists.

List	Nwords in list (max = 445)	Composition
GSL	254	
NGSL_Browne	324	
NGSL_B&G	297	
AWL	16	
Nation	442	L1:305; L2:74; L3:28; L4:12
JACET	396	L1:216; L2:89; L3:40; L4:15
Global scale	401	A1:174; A2:113; B1:68; B2:36; C1:8; C2:2
Profile	424	A1:159; A2:114; B1:92; B2:50; C1:3

Table 4: Distribution of the 445 best-known words in the vocabulary study in the various lists of important words to learn by L2 speakers.

All in all, it is clear that frequency is not a perfect guide to what L2 learners are likely to know. This shows the value of the current test-based approach.

A ranked list of 62 thousand words

In the previous section, we discussed the 445 best known words. We could continue and analyze the complete list, except for the fact that if we do so, we are likely to get lost in detail and swamped by the amount of information. A better approach is to use the list as it is.

One use of the list is to rank the words on difficulty. Two parameters are of use: the percentage of L2 speakers who know the word, and the time L2 speakers need to make a yes-decision. Berger, Crossley, and Kyle (2017) reported that such information is a good predictor of L2 acquisition order, even if the information comes from L1 speakers (participants in the English Lexicon Project; Balota et al., 2007). Therefore, we ranked the words from 1 to 61,851 first on the basis of percentage known, and then on speed of responding for ties. Speed of responding was based on the standardized reaction times of correct yes-responses to words. For this measure, the reaction times are standardized per session, so that big differences in response times between participants do not affect the outcome (for more details, see Mandera et al., 2020).

We can rank the words in the same way for the native speakers (first accuracy, then response speed). This gives us an L1 ranking, which correlates .85 with the L2 ranking, confirming that word recognition in L1 and L2 largely follow the same regularities (Brysbaert, Lagrou, & Stevens, 2017). By looking at the differences between the observed L2 and L1 ranks, we have information about which words are better known by L2 speakers than expected, and which words are less well known. Table 5 illustrates the use by showing the 20 extreme words on each side.

Table 5: Words that are much better known by L2 speakers than expected on the basis of L1 knowledge, and words that are much less well known than expected.

Words better known in L2 than in L1	**Words better known in L1 than in L2**
epicentrum	thumping
paracetamol	drenching
ethnical	eggnog
dismission	rowdy
acceptation	tugging
dynamical	swindle
informatics	sheepishly
unappropriate	clasp
inacceptable	tadpole
addressbook	dunce
energic	pegged
bazar	disgruntled
icecream	runt
recordplayer	grouchy
swimmingpool	pouting
missioner	hoarder
facultative	prancing
analphabet	gnat
policlinic	potbelly
psychical	thimble

From Table 5, it is clear that the L2 speakers in our test have relatively more knowledge of academic words (in particular derived words). Indeed, many of these words are likely to be cognates of words in their L1 or even in another L2. Keuleers, Brysbaert, Stevens, and Mandera (2015) pointed out that the same type of indirect vocabulary likely explains why

Dutch speakers' L1 vocabulary increases with the number of L2s known (see also Aguasvivas et al., 2020, for a similar finding with Spanish L1 speakers). More in general, this suggests that a part of a multilingual's vocabulary can be considered as being *pan-language*. Of course the size of this pan-language lexicon and the ease which it can be transferred between languages is dependent on the specific combination of L1 and L2.

L2 speakers are also more likely to accept morphologically complex words that do not follow English writing conventions. These are compound words written as single words (*addressbook*) and questionable derived forms (*unappropriate, analphabet*).[10]

In contrast, L2 speakers are less familiar with words used in informal social/family settings, illustrating the different uses of English in L1 and L2.

Another analysis we can run to chart the relative strengths and weaknesses of L2 speakers is to make use of the semantic categories defined by Van Overschelde, Rawson, and Dunlosky (2004). These authors distinguished 70 categories and asked participants to type in all the members they could think of in 30 seconds for each category. This resulted in word lists for which we could compare the L1 and L2 ranks. For instance, the category 'carpenter tool' includes (with decreasing frequency of mentioning) *hammer, nail, saw, screwdriver, drill, …, sandpaper, pliers*. Table 6 shows which categories are known by L2 speakers better than expected on the basis of L1 knowledge and which are known worse. Here again we see the influence of differences in language register exposure in L2 and L1 speakers. Words also have a higher chance of being known if they have cognates in various languages. As observed above, many of these words can be considered to be part of a pan-language lexicon.

Table 6: Semantic categories ordered from relatively better known in L2 than expected on the basis of L1 knowledge to known relatively worse in L2. For each category we also give the word with the largest difference in favor of L2 speakers and the word with the largest difference against L2 speakers.

[10] As every other producer of dictionaries, we've been confronted with the frustrating fact that as a list grows in size, it is becoming increasingly difficult not to have a single error in it. Apparently, the comparison of L2 speakers to L1 speakers is a good way to find some of these errors.

Estimated Marginal Means					Best	Worst
			95% Confidence Interval		known	known
Category	Mean	SE	Lower	Upper	in L2	in L2
Toy	-4600	1939	-8405	-795	videogame	doll
Distance	-3810	2168	-8064	443	nanometer	inch
Relative	-3798	1723	-7179	-418	mom	dad
Noisy_thing	-3255	1173	-5556	-953	alarmclock	trumpet
Unit_time	-3016	2083	-7103	1071	millisecond	eon
Science	-2724	2007	-6662	1215	sociology	astronomy
Earth_formation	-2476	1601	-5617	666	rock	glacier
Occupation	-2216	1566	-5289	857	manager	janitor
Reading_material	-2215	1821	-5789	1359	flyer	pamphlet
Music_type	-2015	1723	-5396	1366	rock	folk
Liquid	-1731	1639	-4946	1485	coffee	soda
Non-alcoholic_beverage	-1513	2264	-5956	2930	coffee	soda
Wooden_thing	-1488	1639	-4703	1728	door	bench
Part_of_speech	-1403	2168	-5657	2851	verb	vowel
Sports	-1301	1566	-4374	1771	badminton	lacrosse
Body_part	-1279	1203	-3638	1081	face	torso
Dance	-1251	1821	-4825	2323	tango	waltz
Transportation_vehicle	-1204	1601	-4346	1937	taxi	moped
Color	-1182	1566	-4255	1890	gray	teal
Drug	-1136	1723	-4517	2245	aspirin	pot
Alcoholic_beverage	-909	1939	-4714	2895	vodka	rum
Fruit	-878	1502	-3825	2069	tomato	tangerine
Clothing	-818	1566	-3891	2255	T-shirt	sock
Flying_thing	-810	1723	-4190	2571	superman	moth
Military_title	-783	2007	-4721	3155	captain	admiral
Chemical_element	-693	1445	-3529	2143	uranium	nickel
Green_thing	-329	1679	-3624	2966	stoplight	crayon
Office	-88	2083	-4175	3999	sheriff	mayor
Building_part	-15	1502	-2962	2933	door	stairwell
Fuel	49	1566	-3023	3122	petrol	propane
Metal	72	2007	-3866	4010	gold	brass
Weapon	370	1601	-2772	3512	stick	mace
Four_footed_animal	397	1419	-2388	3181	bear	elk
Building_religion	479	2655	-4731	5688	cathedral	sanctuary
Women_wear	749	1349	-1898	3395	jewelry	thong
Food_flavoring	993	1502	-1954	3940	oregano	nutmeg
Disease	1748	1770	-1725	5221	mononucleosis	smallpox
Furniture	1849	1723	-1531	5230	couch	recliner
Human_dwelling	1870	1601	-1272	5012	box	hut

Weather	1973	1533	-1035	4981	sun	sleet
Vegetable	2882	1566	-190	5955	tomato	turnip
Crime	2962	1878	-722	6646	murder	larceny
Carpenter_tool	3667	1821	93	7240	measurer	pliers
Musical_instrument	3689	1533	681	6697	saxophone	harp
Herb	3973	1821	399	7547	oregano	dill
Ship	4705	1419	1920	7490	sailboat	barge
Insect	4819	1601	1677	7960	butterfly	gnat
Clergy	4969	2007	1031	8907	minister	deacon
Footwear	5113	1939	1308	8918	sandal	clog
Precious_stone	5299	1878	1616	8983	gold	jade
Bird	5328	1395	2592	8064	duck	parakeet
Fabric	5879	1679	2584	9174	nylon	flannel
Kitchen_utensil	6214	1821	2640	9787	plate	ladle
Tree	6426	1566	3354	9499	christmas	spruce
Flower	6800	1878	3116	10484	rose	pansy
Garden_tool	9842	1821	6268	13416	glove	rake
Fish	10231	1502	7284	13178	piranha	trout

A final analysis we performed on the data is how well the L2 and L1 ranks are predicted by word features, such as word length, word frequency, similarity to other words, number of morphemes, age of acquisition, and concreteness. For this analysis, we used the SUBTLEX-US word frequencies from Brysbaert and New (2009), the Orthographic Levenshtein Distances and the number of morphemes from Balota et al. (2007), the age of acquisition ratings from Kuperman, Stadthagen-Gonzalez, and Brysbaert (2012,) and the concreteness ratings from Brysbaert, Warriner, and Kuperman (2014). The data were available for 22,775 words.[11] The outcome of the analysis is shown in Table 7.

Table 7: Importance of variables to predict word ranks in L1 and L2 speakers. Length = word length in number of letters; Freq = log SUBTLEX-US frequency (Brysbaert & New, 2009); OLD = Orthographic Levenshtein Distance to the nearest 20 words (number of letters that must be changed to turn the target word into new, existing English words; Balota et al., 2007); Morph = the number of morphemes in the word (Balota et al., 2007); AoA = age of

[11] Most of these words are generally known words as AoA ratings and concreteness ratings were not collected for words unknown to most native speakers.

acquisition rating (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012); Conc = concreteness rating (Brysbaert, Warriner, & Kuperman, 2014); R² = percentage of variance explained by the models.

Model native speakers	R²	Model bilingual speakers	R²
Length	.006	Length	.004
Length+Freq	.427	Length+Freq	.459
Length+Freq+OLD	.452	Length+Freq+OLD	.472
Length+Freq+OLD+Morph	.459	Length+Freq+OLD+Morph	.498
Length+Freq+OLD+Morph+AoA	.509	Length+Freq+OLD+Morph+AoA	.533
Length+Freq+OLD+Morph+AoA+Conc	.510	Length+Freq+OLD+Morph+AoA+Conc	.533

As can be seen in Table 7, word length and concreteness have negligible impacts on the word ranks, both for L1 and L2 speakers. Most of the variance is explained by word frequency, similarity to other words (OLD), and age of acquisition. Word that are frequent, similar to other words, and acquired early (*believe, mother*) have higher chances of being known than less frequent, late acquired words that look strange (*calipash, gomuti*). Morphologically complex words have a higher chance of being known than expected on the basis of the other variables. The last effect explains more of the variance in L2 speakers than in L1 speakers, in line with what is shown in Table 5. Word frequency also has a stronger impact in L2 speakers than in L1 speakers, as reported by Brysbaert et al. (2017).

A list of 20 thousand word families for learning and teaching

The list of 62 thousand words is interesting to analyze the input for L2 speakers and their output, but is less useful for language teachers. As argued by many researchers, it is better to teach words as part of families than as individual entries. It doesn’t make sense to teach the word forms *are, is, be, were, was, being, been, am* independently. Similarly, when the word form *you* is taught, it makes sense to link it to the forms *your* and *yours*; and when the word *student* is discussed, it seems a lost opportunity not to connect it to *study*. Indeed, L2 learners (often adults) appreciate knowledge of morphological connections more than L1 learners (children). Ullman (2001) hypothesized that children learn language predominantly via procedural memory (implicit memory for actions) whereas adults learn language predominantly via declarative memory (explicit memory for facts and relations). Later research showed that L2 acquisition makes more use of implicit learning than postulated by Ullman, but kept on observing that explicit knowledge about rules and word relationships plays a larger role in L2 acquisition than in L1 acquisition (Williams & Rebuschat, 2017).

Language teachers can use the list of 62 thousand words to make their own levels for word families. For instance, they can use Nation's list of 25 thousand word families. Alternatively, they can make use of the list of 18 thousand word families provided by Brysbaert et al. (2016). In the supplementary materials, we provide a list of 20 thousand word families obtained by summing all inflections and derivations based on suffixes (but not prefixes). So, the word family of *correction* (the best known word of the family) includes *correction, corrected, correct, correctly, correcting, corrective, corrector, correctional, correctable, correctness, correctible*, and *correctively*, but not *incorrect*. The word families are divided in 20 levels of 1,000 families each. The levels are based on the rank of the best known family member. In this way, the teaching order mimics the L2 learners' "typical" acquisition order. An advantage of this ranking is that the teaching does not start with all function words (which happens when the order within the levels is based on frequency). A disadvantage is that some function words appear rather late, because people regularly failed to press "yes" to function words (e.g., 2% errors to the word *of*). So, occasionally teachers will feel a need to deviate from the list and it is better to see the ranks as guidelines rather than a strict framework. Similarly, there is no need to teach all members in a family. Much better to teach only those that are known to more than 95% of L2 speakers.

Discussion

Because of the multitude of words L2 learners must acquire in their new language, researchers have made various lists with levels of word usefulness, so that language teachers can start with the most useful words. Frequency of use has been the main criterion to compile the lists. Our research shows that this indeed is a useful criterion but only accounts for 46% of the variance in the likelihood that a word will be known to an L2 speaker.

There are at least three reasons that limit the use of word frequency norms to predict L2 word knowledge. The first is that the corpus on which the word frequencies have been calculated is not entirely representative for the language input L2 language learners receive. Speech in films and television series (on which the SUBTLEX word frequencies are based) may form one source of input for L2 learners, but is unlikely to be sole source or the most important. Ideally, we would be able to compile a corpus consisting of all input groups of L2 learners are exposed to.

A second reason why word frequency is a limited predictor of L2 word knowledge is that word frequency is based on the assumption that all encounters with words have the same weight. Words encountered more often are known better, independent of other word characteristics. This is true to some extent, but not completely. Some words seem to be more difficult to learn than other, depending on the concept they represent, the similarity to other words (both in terms of meaning and form) and, in the case of L2 words, the similarity to L1 translations (He & Godfroid, 2019; Laufer & Nation, 2012).

Researchers have tried to improve the predictability of word knowledge by including more word-related variables than word frequency alone, such as word length, orthographic similarity to other words, concreteness, semantic richness, valence, and so on. A good example is Hashimoto and Egbert (2019), who presented a list of 500 English words to 403 nonnative adult English language learners with many different L1 backgrounds. The vocabulary test consisted of yes/no decision like in the present article. The authors used a range of word-related variables to predict how many participants would know a word. The best predictors were (1) the McDonald co-occurrence probability, which measures the likelihood of a word statistically co-occurring (within five tokens to either the left or the right) with 500 highly frequent lemmas in the British National Corpus, (2) word frequency (based on the Corpus of Contemporary American English), (3) number of senses, and (4) the number of words in the orthographic neighborhood with frequency greater than the item's frequency. Together the variables accounted for 37% of the variance in word knowledge. Word frequency alone accounted for 25% of the variance.

A third reason why word frequency is a limited predictor is that word learning not only depends on the characteristics of the word but also on the motivation of the learner to know the word. As Laufer and Nation (2012, p. 164) remarked: “some words may be useful to some learners even if they are not generally frequent in the language, depending on the specific needs that learners may have”. Why do nearly all L2 learners know words such as *midnight, snowboarding*, and *sexy*? Arguably not only because of form- and frequency-related aspects of these words, but also because the words are perceived as “interesting to know” (i.e., as useful to the user).

If perceived interest/usefulness contributes to word knowledge (Keuleers, 2018), we can expect that our measure of word learning rank predicts extra variance in the data of Hashimoto and Egbert (2019), other than the word-form related variables the authors used. Indeed, the correlation between word knowledge in Hashimoto and Egbert (2019) and the L2 ranks from the present study is .68 (i.e., 46% shared variance), even though we use only one predictor.[12] Interestingly, the correlation between word knowledge in Hashimoto and Egbert and the L1 ranks from the present study is only .28 (8% shared variance), confirming that word learning in L2 differs from word learning in L1, certainly for the first learned words (see Berger et al., 2017, for a study predicting L2 text processing cost on the basis of L1 word difficulty).

Another interesting dataset was published by He and Godfroid (2019). These authors collected measures of word frequency, word usefulness and word difficulty for 191 English academic words and multiword expressions. Word frequency was based on counts in the academic texts of COCA (Gardner & Davies, 2014). Word usefulness was estimated by asking English L2 teachers how worthwhile the words were to teach to students preparing for college in North America (1 = not worth, 7 = indispensable). Word difficulty was rated by the same teachers who were asked to indicate how advanced the vocabulary was (1 = basic/easy, 7 = advanced/difficult). The present article includes ranks for 164 of the 191 stimuli (the other stimuli were multiword expressions and the word “contaminated”). The L2 ranks correlated most with word difficulty ($r = .60$), then with log frequency ($r = -.52$) and with word usefulness ($r = -.47$). Multiple regression indicated that all three variables contributed significantly to the prediction of L2 ranks ($R^2 = .46$). Correlations with the L1

[12] We thank the authors for kindly providing the data to us.

ranks were much lower and only word difficulty contributed significantly in a multiple regression analysis ($R^2 = .19$).

All in all, the L2 ranks obtained on the basis of word prevalence and word recognition times in a crowdsourcing study form an interesting variable reflecting (1) how difficult a word is to learn, (2) how frequently an L2 speaker is likely to come across it, and (3) how motivated the speaker is to learn the word (He & Godfroid, 2019). In addition, the variable generalizes well to other studies (Hashimoto & Egbert, 2019). L2 word knowledge is interesting to optimize vocabulary teaching[13] and for research purposes. We think it will help researchers to better estimate text difficulty (Berger et al., 2017) and to match stimuli when the influence of new variables is investigated. The ranks are also interesting as dependent variable because they allow researchers to examine which variables affect word acquisition (in addition to word difficulty, frequency and perceived usefulness).

At the same time, it is important to keep in mind the limitations of our dataset. A first limitation is that our lists are based on yes/no word decision. Although this test has been validated several times (Ferré & Brysbaert, 2017; Harrington, & Carey, 2009; Lemhöfer & Broersma, 2012; Meara, & Buxton, 1987; Zhang et al., 2019), it remains true that it does not measure word knowledge in great depth. Participants may know that a letter sequence is an English word, without much knowledge about the meaning of the word, let alone the various senses and unrelated meanings English words may have. Also, the yes/no task seems to underestimate the knowledge of function words, possibly because these words rarely occur in isolation or because participants do not expect such words in a vocabulary test.

The limitation of the yes/no decision can be seen in the surprisingly high number of words "known" by speakers who indicate they speak but a few words of English (21 thousand words; see Table 1), although several factors are involved in addition to hazy word knowledge. First, some of this performance is due to guessing (with a maximum of 7% given that we excluded sessions with more than two yes-responses to non-words). Second, it is likely that many of the people with little knowledge were too humble about their level, given

[13] As a reviewer remarked, from a curriculum design viewpoint word prevalence may be viewed as a complement to word frequency (rather than as a competitor). Prevalence then represents what is already known to students of a certain proficiency level, whereas frequency indicates which new words are most interesting to teach.

the small difference between the categories “know a few words” and “am able to hold a simple conversation”. In hindsight, it is improbable that participants without considerable knowledge of English were able to answer the person-related questions (which was a condition to be included in the analyses). Third, the words refer to a smaller number of word families, which is the currency of existing estimates (Laufer & Ravenhorst-Kalovski, 2010); Nation, 2006). Finally, many languages have a considerably number of English cognates (e.g., scientific words) and loan words. This is particularly true for languages using the Roman alphabet (the bulk of our participants). Also, in many countries exposure to English is high. For instance, De Wilde, Brysbaert, and Eyckmans (2020) observed that 25% of the Dutch-speaking children in Belgium understand English at the A2-level at the end of primary school, even before they get any formal education of the language.

Still, it will be interesting to see how the present results compare to more demanding word knowledge tasks, such as giving a synonym or translation.[14] In this respect, it may also be true that some of our non-words were too easy. These were derived from all English words, irrespective of their frequency and the knowledge of them (i.e., words like *yapok* and *clypeal* had the same weight as *happy* and *music*). As a result, some of the non-words looked rather un-English, which may have made it tempting for participants to say “yes” to English-looking words they did not know. A further discussion of how to create good non-words can be found in König, Calude, and Coxhead (2019).

Another limitation of our study is that the list contains words only. An important part of a language consists of multiword expressions, such a phrasal verbs (*give up, give in*), compound words (*washing machine*), idioms (*call it a day*) and other types of formulaic language (e.g, *excuse me*). Knowledge of multiword expressions is an important part of language proficiency and is the logical next research step. An initial collection of such expressions suggests that the list in English is at least as long as that of single words (i.e., over 60 thousand multiword expressions included in various dictionaries and collections).

Despite the limitations we are confident that the lists collected in the present research will be of much use to English L2 researchers and teachers, because for the first time we get

[14] Norbert Schmitt (personal communication) is collecting such data.

extensive information about the level of knowledge for nearly all words in the English language.

Availability

The lists that have been made for the present article are available as Excel files in supplementary materials and on https://osf.io/gakre/. They include:

- The list of 114 words known to all L2 participants (Table 3).
- The list of 445 words with zero or one mistake.
- The list of 61,851 words with accuracy rates, response times, and ranks (both for L2 and L1).
- The list of semantic category exemplars with ranks both in L1 and L2 (Table 6)
- The list of 22,775 words included in the regression analyses (Table 7), together with the predictor variables.
- The list of 20,000 word families for teaching and the levels they have been assigned to.

References

Aguasvivas, J., Carreiras, M., Brysbaert, M., Mandera, P., Keuleers, E., & Dunabeitia, J.A. (2020). How do Spanish speakers read words? Insights from a crowdsourced lexical decision megastudy. *Behavior Research Methods*. Preprint available at https://doi.org/10.3758/s13428-020-01357-9

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39(3)*, 445-459.

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography, 6(4),* 253-279.

Benigno, V., & de Jong, J. (2019). Linking vocabulary to the CEFR and the Global Scale of English: A psychometric model. In A. Huhta, G. Erickson, & N. Figueras (Eds), *Developments in Language Education: A Memorial Volume in Honour of Sauli Takala* (pp. 8-29). University Printing House, Jyväskylä, Finland.

Berger, C. M., Crossley, S. A., & Kyle, K. (2017). Using Native-Speaker Psycholinguistic Norms to Predict Lexical Proficiency and Development in Second-Language Production. *Applied Linguistics, 40(1),* 22-42.

Brezina, V., & Gablasova, D. (2013). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics, 36(1)*, 1-22.

Browne, C. (2014). A new general service list: The better mousetrap we've been looking for. *Vocabulary Learning and Instruction, 3(2)*, 1-10.

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition, 20*, 530-548.

Brysbaert, M., Keuleers, E., & Mandera, P. (2019a). Recognition Times for 54 Thousand Dutch Words: Data from the Dutch Crowdsourcing Project. *Psychologica Belgica, 59*, 281–300.

Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019b). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods, 51(2)*, 467-479.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977-990.

Brysbaert M, Stevens M, Mandera P and Keuleers E (2016) How Many Words Do We Know? Practical Estimates of Vocabulary Size Dependent on Word Definition, the Degree of Language Input and the Participant's Age. *Frontiers in Psychology, 7:1116*. doi: 10.3389/fpsyg.2016.01116

Brysbaert, M., Warriner, A.B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*, 904-911.

Capel, A. (2012). Completing the English vocabulary profile: C1 and C2 vocabulary. *English Profile Journal, 3*. DOI: https://doi.org/10.1017/S2041536212000013

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology 11(3)*, 38-63.

Cobb, T. (2016) Numbers or Numerology? A response to Nation (2014) and McQuillan (2016) *Reading in a Foreign Language, 28 (2),* 299-304.

Coxhead, A. (2000). A new academic word list. TESOL Quarterly, 34(2), 213-238.

Dang, T. N. Y., Webb, S., & Coxhead, A. (2020). Evaluating lists of high-frequency words: Teachers' and learners' perspectives. *Language Teaching Research*. Preprint available at https://doi.org/10.1177/1362168820911189.

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2020). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language & Cognition, 23*, 171-185. https://doi.org/10.1017/S1366728918001062.

Dürlich, L., & François, T. (2018, May). EFLLex: A Graded Lexical Resource for Learners of English as a Foreign Language. In *Proceedings of the Eleventh International Conference on*

Language Resources and Evaluation (LREC-2018). https://www.aclweb.org/anthology/L18-1140.

Ferré, P., & Brysbaert, M. (2017). Can Lextale-Esp discriminate between groups of highly proficient Catalan-Spanish bilinguals with different language dominances? *Behavior Research Methods, 49*, 717-723.

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics, 35(3)*, 305-327.

Gineste, R., & Lagrave, R. (1961). *Le français fondamental par l'action [Fundamental French by action]*. Edition Didier.

Harrington, M., & Carey, M. (2009). The on-line Yes/No test as a placement tool. *System, 37(4)*, 614-626.

Hashimoto, B. J., & Egbert, J. (2019). More than frequency? Exploring predictors of word difficulty for second language learners. *Language Learning, 69(4)*, 839-872.

He, X., & Godfroid, A. (2019). Choosing Words to Teach: A Novel Method for Vocabulary Selection and Its Practical Application. *TESOL Quarterly, 53(2)*, 348-371.

Keuleers, E. (2018). *A transaction perspective on the language environment* (Plenary). Paper presented at 4th Usage Based Linguistics Conference, Tel Aviv, Israel.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42*, 627-633.

Keuleers, M., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology, 68*, 1665-1692.

König, J., Calude, A. S., & Coxhead, A. (2019). Using Character-Grams to Automatically Generate Pseudowords and How to Evaluate Them. *Applied Linguistics.* Advance publication at https://doi.org/10.1093/applin/amz045.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. Behavior Research Methods, 44, 978-990.

Laufer, B., & Nation, I. S. P. (2012). Vocabulary. In S. M. Gass & A. Mackey(Eds.), *The Routledge handbook of second language acquisition* (pp. 163–176). London, England: Routledge.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a Foreign Language, 22(1)*, 15-30.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. Behavior Research Methods, 44(2), 325-343.

Mandera, P., Keuleers, E., & Brysbaert, M. (2020). Recognition times for 62 thousand English words: Data from the English Crowdsourcing Project. *Behavior Research Methods, 52(2)*, 741-760.

Meara, P. M., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing, 4*, 142–154.

Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63(1)*, 59-82.

Nation, I.S.P., & Waring, R. (1997). Vocabulary size, text coverage and words lists. In N. Schmitt & M. McCarthy (Eds.): *Vocabulary: Description, Acquisition and Pedagogy* (pp. 6-19). Cambridge: Cambridge University Press.

Uemura, T., & Ishikawa, S. I. (2004). JACET 8000 and Asia TEFL vocabulary initiative. *Journal of Asia TEFL, 1(1)*, 333-347.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and cognition, 4(2)*, 105-122.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the norms. *Journal of Memory and Language, 50(3),* 289-335.

Webb, S. A., & Chang, A. C. S. (2012). Second language vocabulary growth. *RELC Journal, 43(1)*, 113-126.

West, M. (1953). *A general service list of English words*. London: Longman, Green.

Williams, J., & Rebuschat, P. (2017). *Implicit Learning and Second Language Acquisition.* Routledge.

Zhang, X., Liu, J., & Ai, H. (2019). Pseudowords and guessing in the Yes/No format vocabulary test. *Language Testing*. Advance publication on https://doi.org/10.1177/0265532219862265.