



LAB: Linguistic Annotated Bibliography – A searchable portal for normed database information

Journal:	<i>Quarterly Journal of Experimental Psychology</i>
Manuscript ID:	Draft
Manuscript Type:	Special Issue Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Buchanan, Erin; Missouri State University, Psychology Carr, Michael; Missouri State University, Psychology Valentine, Kathrene; Missouri State University, Psychology Bankovich, Farren; Missouri State University, Psychology Saxton, Samantha; Missouri State University, Psychology Segui, Emmanuel; Missouri State University, Psychology
Keywords:	database, stimuli, online portal, megastudy, trends

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Running head: LINGUISTIC BIBLIOGRAPHY

1

LAB: Linguistic Annotated Bibliography – A searchable portal for normed database information

Erin M. Buchanan

Michael T. Carr

Kathrene D. Valentine

Farren E. Bankovich

Samantha D. Saxton

Emmanuel Segui

LINGUISTIC BIBLIOGRAPHY

2

Abstract

In the era of big data, psycholinguistic research is flourishing with numerous publications advancing our knowledge of word characteristics and ways to study them. This article presents the Linguistic Annotated Bibliography (LAB) as a searchable web portal to quickly and easily access reliable database norms, related programs, and variable calculations. These publications ($N = 561$) were coded by language, number of stimuli, stimuli type (i.e. words, pictures, symbols), keywords (i.e. frequency, semantics, valence), and other useful information. This tool not only allows researchers to search for the specific type of stimuli needed for experiments, but also permits the exploration of publication trends across 100 years of research. Details about the portal creation and use are outlined, as well as various analyses of change in publication rates and keywords. In general, advances in computation power have allowed for the increase in dataset size in the recent decades, in addition to an increase in the number of linguistic variables provided in each publication.

Keywords: *database, stimuli, online portal, megastudy, trends*

LINGUISTIC BIBLIOGRAPHY

LAB: Linguistic Annotated Bibliography – A searchable portal for normed database information

The advance of computational ability and the Internet have propelled research into an era of so-called *big data* that have interesting implications for the field of psycholinguistics, as well as other experimental areas that use normed stimuli for their research. Traditionally, stimuli used for experimental research in psycholinguistics were first examined through small pilot studies that were then used in many subsequent projects. While economic, that selection procedure’s results could be potentially misleading as a factor of the stimuli, rather than experimental manipulation. This limitation can potentially be tied to a lack of funding, time, computing power, or even interest in studying phenomena at the stimuli level. Now, we have the capability to collect, analyze, and publish large datasets for research into memory models (Cree, McRae, & McNorgan, 1999; Moss, Tyler, & Devlin, 2002; Rogers & McClelland, 2004; Vigliocco, Vinson, Lewis, & Garrett, 2004), aphasia (Vinson, Vigliocco, Cappa, & Siri, 2003), probability and linguistics (Cree & McRae, 2003; McRae, de Sa, & Seidenberg, 1997; Pexman, Holyk, & Monfils, 2003), valence (Dodds, Harris, Kloumann, Bliss, & Danforth, 2011; Vo et al., 2009; Warriner, Kuperman, & Brysbaert, 2013), and reading speeds and priming (Balota et al., 2007; Cohen-Shikora, Balota, Kapur, & Yap, 2013; Hutchison et al., 2013; Keuleers, Lacey, Rastle, & Brysbaert, 2012) to name a small subset of research avenues.

Big data has manifested in psycholinguistics over the last decade in the form of grant funded megastudies to collect and analyze large text corpora (the SUBTLEX projects) or to examine numerous word properties in one study (the Lexicon projects). The SUBTLEX projects were designed to analyze frequency counts for concepts across extremely large corpora sizes using subtitles as a substitute for natural speech. The investigation of these measures was first spurred by the realization that word frequency is an important predictor of naming and lexical

LINGUISTIC BIBLIOGRAPHY

4

1
2
3 decision times (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Rayner & Duffy,
4
5 1986). While previous measures of frequency (i.e. Baayen, Piepenbrock, & Gulikers, 1995;
6
7 Burgess & Livesay, 1998; Kucera & Francis, 1967) were based on large 1 million + word
8
9 corpora, they were poor predictors of reaction times (Balota et al., 2004; Brysbaert & New,
10
11 2009; Zevin & Seidenberg, 2002). Further, it appears from Brysbaert and New's (2009)
12
13 investigation into corpora size and type, that not only should the corpora be large (>16 million),
14
15 but the underlying source of the text data matters (Internet versus subtitles), as well as the
16
17 contextual diversity of the data (i.e. number of occurrences across sources; Adelman, Brown, &
18
19 Quesada, 2006). Not only has their work (Brysbaert & New, 2009) been included in newer
20
21 lexical studies (Hutchison et al., 2013; Yap, Tan, Pexman, & Hargreaves, 2011), but SUBTLEX
22
23 projects have been published in Dutch (Keuleers, Brysbaert, & New, 2010), Greek
24
25 (Dimitropoulou, Dunabeitia, Aviles, Corral, & Carreiras, 2010), Spanish (Cuetos, Glez-Nosti,
26
27 Barbon, & Brysbaert, 2011), Chinese (Cai & Brysbaert, 2010), French (New, Brysbaert, Veronis,
28
29 & Pallier, 2007), British English (van Heuven, Mandera, Keuleers, & Brysbaert, 2013), and
30
31 German (Brysbaert, Buchmeier, Conrad, Jacobs, Bolte, & Bohl, 2011).

32
33
34
35
36
37
38
39 The Lexicon projects created large databases of validated mono- and multisyllabic words
40
41 to assist in the creation of controlled experimental stimuli sets for future experiments. These
42
43 databases contain lexical decision and naming response times, as well as typical word confound
44
45 variables such as orthographic neighborhood, phonological and morphological characteristics.
46
47 While the English Lexicon Project (Balota et al., 2007) is the most cited of the lexicons, other
48
49 languages include Chinese (Sze, Rickard Liow, & Yap, 2013), Malay (Yap, Liow, Jalil, &
50
51 Faizal, 2010), Dutch (Keuleers, Diependaele, & Brysbaert, 2010), and British English (Keuleers,
52
53 Lacey, Rastle, & Brysbaert, 2012). Another twenty or so similar lexical database publications
54
55
56
57
58
59
60

LINGUISTIC BIBLIOGRAPHY

5

1
2
3 can be found in the literature covering French (Lete, Sprenger-Charolles, & Cole, 2004), Italian
4
5 (Barca, Burani, & Arduino, 2002), Arabic (Boudelaa & Marslen-Wilson, 2010), and Portuguese
6
7 (Soares et al., 2013).
8
9

10 The availability of big data has augmented the psycholinguistic literature, but these
11
12 projects are certainly time consuming due to the amount of participant data required to achieve
13
14 reliable and stable norms. The solution potentially lies in several avenues of easily obtainable
15
16 data. First, Amazon's Mechanical Turk, an online crowdsourcing avenue that allows researchers
17
18 to pay users to complete questionnaires, has shown to be a reliable, diverse participant pool made
19
20 available at very low cost (Buhrmester, Kwang, & Gosling, 2011; Mason & Suri, 2012).
21
22 Researchers can pre-screen for specific populations, as well as post-screen surveys for
23
24 incomplete or inappropriate responses, thus saving time and money with the elimination of poor
25
26 data. Because of the popularity of Mechanical Turk, large amounts of data can be collected in
27
28 shorter time periods than traditional experiments. Mechanical Turk has been used to collect data
29
30 for semantic word pair norms (Buchanan, Holmes, Teasley, & Hutchison, 2013), age of
31
32 acquisition ratings (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012), concreteness ratings
33
34 (Brysbaert, Warriner, & Kuperman, 2013), past tense information (Cohen-Shikora, et al., 2013),
35
36 and valence and arousal ratings (Dodds, Harris, Kloumann, Bliss, & Danforth, 2011; Jasmin &
37
38 Casasanto, 2012; Warriner et al., 2013). Additionally, in a similar vein to the SUBTLEX
39
40 projects, linguistic data has been mined from open source data, such as the New York Times,
41
42 music lyrics, and Twitter (Dodds et al., 2011; Kloumann, Danforth, Harris, Bliss & Dodds,
43
44 2012). Finally, De Deyne, Navarro, and Storms (2013) have seen success in simply setting up a
45
46 special website (www.smallworldofwords.com) to collect word pair association norms.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

LINGUISTIC BIBLIOGRAPHY

6

The evolution of big data provides exciting opportunities for exploration into psycholinguistics, and this article features the trends in publications of normed datasets across the literature allowing for a large-scale picture of the developments of trends in psychological stimuli. Historically, these norms have been published in journals connected to the *Psychonomic Society*, such as *Behavior Research Methods*, *Psychonomic Monograph Supplements*, and *Perception and Psychophysics*. The society once hosted an electronic database that contained the links to these norms, as well as a search tool to find information about previously published works (Vaughan, 2004). The sale of the society journals to Springer publications has improved journal visibility and user-friendly access, but also has left a need for an indexed list of database publications that span multiple keywords and journal websites. Therefore, the purpose of this article is twofold: 1) to present a searchable, cataloged database of normed stimuli and related materials for a wide range of experimental research, and 2) to examine trends in the publications of these articles to assess the big data movement within psycholinguistics.

Website

Readers can find the website by going to the host site for semantic word pair norms (Buchanan et al., 2013) at <http://www.wordnorms.com> or by going to the direct link at <http://wordnorms.missouristate.edu>. The wordnorms.com website will redirect you to the direct link and was meant to be an easier web address to remember, in addition to a permanent address in case of a change in hosting university. From this page, the top navigation bar includes a link for *Norms* to direct the reader to the LAB page. The main LAB page, as shown in Figure 1 includes the purpose statement, and several website options. From here, users can suggest articles that should be included in the dataset by clicking on *Enter data here*, view all the data in an easy to copy format (*View all data big*) or in a smaller more readable format (*View all data*

small), search the database with large, easy to copy table (*Search with large data output*) or small table formats (*Search with small data output*) and view many of the tables presented here. These tables are dynamic, and they update with each addition to the database, which allows the user to view current statistics even after publication. Although the website was designed to be intuitively user friendly, a complete how-to guide is included online for unfamiliar users. Specific features will be outlined below in relation to the database creation.

Database

Method

Materials

Bradshaw (1984) and Proctor and Vu’s (1999) lists of database information were used as starting points for collection of research articles. We searched Academic Search Premier, PsycInfo, and ERIC through the EBSCO host system, as well as Google scholar to find other relevant articles using the following keywords: *corpus*, *linguistic database*, *linguistic norms*, *norms*, and *database*. Additionally, since many of the original articles were hosted by the Psychonomic Society, the Springer website was searched with these terms that covered the newer editions of *Behavior Research Methods* and *Memory & Cognition*. We then filtered for articles that met the following criteria: 1) contained database information as supplemental material, 2) demonstrated programs related to building research stimuli using normed databases, or 3) generated new calculations of lexical variables. Research articles that used normed databases in experimental design or tested those variables validity/reliability were excluded if they did not include new database information. Additional articles were found while coding initial publications by searching citations for stimuli selection. For example, the Snodgrass and Vanderwart (1980) norms were cited in many newer articles on line drawings, and therefore this

LINGUISTIC BIBLIOGRAPHY

8

article was subsequently entered into the database. At the time of writing, 561 articles, books, websites and technical reports were included in the following analyses.

Coding Procedure

The tables with summaries from Bradshaw (1984) and Proctor and Vu (1999) were consulting for a starting point for data coding. Then, the first round of articles found (approximately 100) from the methods described above were analyzed to determine information that would be pertinent to a user who wished to search for normed stimuli. Based on these reviews and lab discussions, we coded the following information from each article: 1) journal information, 2) stimuli types, 3) stimuli language, 4) programs or corpus name, 5) keywords, which we refer to as tags, 6) special populations, and 7) other notes that did not fit into those categories. Each piece of information is detailed below. In some instances, codes were not used as frequently as expected based on these initial discussions, but were included to allow more specificity in searching, as well as the flexibility to include those options for articles subsequently added to the database.

Journal Information. Each article was coded with the citation information, and a complete list of citations can be found on the website portal by clicking on view all data. All author last names are listed, along with publication year, article title, journal title, volume, page numbers, and digital object identifier (DOI) when available. This information is listed in citation format in the small table output, and separated into columns in the large table output for easier sorting and searching. For newer articles that have been published online first without volume or page numbers, *X* and *XX-XX* are used as placeholders until official publication. DOIs in all table outputs are hyperlinked to the article on the publication journal's website, which is accessible if the user has access through their membership with a professional organization or university.

LINGUISTIC BIBLIOGRAPHY

Although APA style dictates et al. for references after the second author or immediately for large author publications, all names were included each time they were referenced (see below). The inclusion of these names allows a user to search for specific researchers, as well as separates different publications by the same first author. A complete list of publication sources, number of times cited, and percentages can be found online by clicking Journal Frequency Table.

Stimuli Types. While this publication was originally intended for linguistic database norms, other types of experimental stimuli were apparent after background review. Therefore, stimuli were coded based on the dominant description from the article (i.e. although heteronyms are words and word pairs, they were coded specifically as heteronyms). The number of stimuli presented in the appendix or database was coded with the stimuli, unless the article covered specific programs, search or experimental creation tools (the majority of the *other* category). Because many articles included two types of stimuli, or references to different articles where stimuli were selected from, two options for stimuli were included. Therefore, the total values for number of stimuli do not add up to the number of articles in the database because of multiple instances in articles. Table 1 includes a stimuli list, the number of times that each stimuli was used, percentage of the total stimuli codes, the mean and standard deviation of the number of those stimuli, minimum/maximum values, and a brief variable description. Researchers often cited specific previous works where stimuli were selected from, and these references were included in the stimuli column. Further, if stimuli were associated with a specific database (such as *CELEX*: Baayen, Piepenbrock, & Gulikers, 1995 and *ANEW*: Bradley & Lang, 1999), those abbreviations were included for searching capabilities. Table 1 is included dynamically online, such that updates are included automatically, and can be found under the Stimuli Frequency Table link.

LINGUISTIC BIBLIOGRAPHY

10

Stimuli Language. The language of the stimuli set was coded by starting with the most common languages from the first articles surveyed, and others were added as it was apparent that several norms were present for that language (such as Japanese, Dutch, and Greek). If the stimuli were non-word selections, like pictures and line drawings, the language of the participants used to norm the set was used, which was commonly English. The *other* category was used for low-frequency languages, as well as a multiple category for datasets with more than one set of language norms. One potential limitation of the LAB was that English is the first language for the authors; however, translation tools were used to code sources found in other languages. The LAB portal includes options to report errors in coding, as well as a form to enter new articles that may have been missed due to this drawback. Table 2 shows language frequencies and percentages, and the online version can be found by clicking the Language Frequency Table link.

Program/corpus name. In many instances, megastudies are often named, such as the English Lexicon Project (Balota et al., 2007), for easier reference. This information was included in the dataset, which will also help researchers with the stimuli references as described above. For example, a newer study may reference using the BOSS database (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010), and having that information would make searching for the original article easier by using the corpus name column (especially in instances the dataset name is not listed in the article title). The names of programs or tools were also entered, such as NIM (Guasch, Boada, Ferre, & Sanchwz-Casas, 2013), a new stimuli selection tool for psycholinguistic studies.

Tags. Keyword tags are the majority of the database, as they allow for the best understanding of trends and availability of stimuli. Table 3 shows a list of tags, frequencies, percentages, descriptions, and correlations (described below). Each article was coded with tags based on the

LINGUISTIC BIBLIOGRAPHY

11

1
2
3 description of the accessible data, so that one article may have many tags. However, due to the
4
5 cumulative nature of database research, this tagging system does not mean that each article
6
7 collected that particular type of data. The most common example of this distinction occurs when
8
9 data is combined across sources, but presented in a new article. The Maki, McKinley, and
10
11 Thompson (2004) semantic distance norms also included values from the South Florida Free
12
13 Association norms (Nelson, McEvoy, Schreiber, 2004), and Latent Semantic Analysis (Landauer
14
15 & Dumais, 1997). Therefore, this article was coded with association and semantics, even though
16
17 the association norms were not collected in that paper. As described above, some small
18
19 frequency tags were used because of the initial pass through newer articles, but these were left in
20
21 the database because of their specificity, and they can be used in future additions.
22
23
24
25
26

27 **Special Populations.** While coding articles, it became apparent that a subset of the normed data
28
29 was tested on specific special populations. Consequently, demographic data such as gender, age,
30
31 ethnicity, and grade school year were listed as described in the article (i.e. if ages were used, age
32
33 was listed, but if grade year was used, it was listed rather than translating to specific ages).
34
35

36 **Other/Notes.** Lastly, places for more description were included for tags or variables not
37
38 frequently used, which was especially useful for program descriptions, as well as descriptions of
39
40 specific types of stimuli (i.e. CVC trigrams). In several instances, notes that appeared frequently
41
42 were moved to tags (such as similarity) after the database had several hundred articles sampled.
43
44 All information described above without a specific table (special populations, other,
45
46 program/corpus names, and journal information) can be found by clicking on either the small or
47
48 large view data links online.
49
50
51
52

53 **Results and Discussion**
54
55
56
57
58
59
60

LINGUISTIC BIBLIOGRAPHY

12

Journals. Journal results, unsurprisingly, show that the wealth of data was published in *Behavior Research Methods* (59.42% combined across name changes). However, a large number of articles also appeared in *Psychonomic Monograph Supplements* (3.41%), *Journal of Verbal Learning and Verbal Behavior* (2.69%), *Psychonomic Science* (2.69%), *Journal of Experimental Psychology* (combined across subjournals, 2.34%), *Perception & Psychophysics* (2.33%), *Memory & Cognition* (2.15%), *Bulletin of the Psychonomic Society* (1.26%), *Norms of Word Association* (1.44%; Postman & Keppel, 1970), and a large category of other publications (21.24%). The complete list can be found in the Journal Frequency Table online, as there were 107 different entries for journals, books, and websites of publications. While some of these sources were not published with peer review (i.e. websites), they were generally found through citations of other peer-reviewed work. Although *Behavior Research Methods* has dominated the field for publications, the large array of options for publishing indicates a growth in the available avenues for researchers in this field (for example, open source journals such as *PLoS ONE* and websites).

Figure 2 portrays the number of publications in half decade intervals, and there has been a clear expansion of database and program papers, as part of the growth in big data. Interestingly, a first growth of publications tracks with the 1950s cognitive revolution (Miller, 2003), but an odd decline in publications occurred from the 1970s to 1990s. The last twenty years has shown unbelievable progress in this area, at over 121 publications in the last four years alone (2010-2013). This chart can be found in greater detail online, under the Papers Per Year Graph link, showing the ups and downs of publications by year in a larger format. For example, 2004, 2010 and 2013 were big years for linguistic publications, while 2005 and 2006 had smaller numbers of

LINGUISTIC BIBLIOGRAPHY

publications. Even with these fluctuations, a clear growth curve in publications can be found since the 90s.

Stimuli. Stimuli are presented in Table 1 with totals, percentages, means (standard deviations), and minimum/maximum values (and online under Stimuli Frequency Table). Naturally, the publication of word norms was over half the dataset (50.55%), which has quite a large range of quantity of stimuli from only ten words to a large corpus of over 500 million words. Other types of word stimuli also appear commonly in the stimuli set such as categories, letters, and word pairs. Because linguistic data was of interest to this publication, we selected publications based on words, and plotted the number of stimuli presented in the paper to examine big data developments. These data were broken down by set size in Figure 3. The upper left hand quadrant shows all stimuli across years, and the big data publications stand out in the last fifteen years of publications. This data was then further broken down into small datasets (<1,000 stimuli; upper right quadrant), medium datasets (1,000 – 500,000 stimuli; bottom left quadrant), and large datasets (500,000 + although there is a large jump between medium and large as most data is either half million or less or a million or more; bottom right quadrant). The small dataset graph shows that these publications are common across time, while the bottom two quadrants are more telling for the megastudies trend investigation. As with languages and tags (below), we see an increase in the number of medium and very large datasets across the years (the lone outlier in the large dataset is the Brown Corpus, Kucera & Francis, 1967).

Languages. The variety and number of languages for stimuli provided an encouraging picture of the growth and diversity of psycholinguistic stimuli, as seen in Table 2. Many articles include multiple languages (4.10%), as well as the inclusion of both Portuguese (1.78%) and Spanish (6.60%), French (5.70%), and even a small number of articles in Greek (0.89%). To examine

LINGUISTIC BIBLIOGRAPHY

14

1
2
3 trend, the English only articles were filtered out of the dataset since they were the majority of
4
5 publications (63.99%), and were published across all years present in this data. More than half of
6
7 the papers with multiple languages have been published since 2010 (56.52%). Additionally, the
8
9 last ten years have seen an explosion of publications in non-English languages (134 publications,
10
11 74.81%) with nearly double the number in 2013 (16) than 2012 (9).

12
13 **Tags.** Table 3 displays the number, percentages, correlations, and descriptions of tags (also
14
15 found online under Tag Frequency Table with N and % values based on any updated numbers).
16
17 Undoubtedly, these tags represent changes in terminology over time, and some could be
18
19 combined or recoinced. However, even if low frequency ($N < 10$; fourteen tags) tags are
20
21 excluded, thirty-five different tags were used to describe the types of psycholinguistic data.
22
23 Many of these tags can be considered individual research areas, and the sizeable number of
24
25 different options indicates how complex and diverse the field has become since the publication
26
27 of free association norms in 1910 (Kent & Rosanoff association norms).
28
29
30
31
32
33

34 The total number of tags for each publication was then tallied, and an average number of
35
36 tags per half decade were plotted in Figure 4 to determine if the number of variables included in
37
38 a study has grown over time (total tags and year $r = 0.26$, $p = 0.10$). Considering the larger
39
40 number of publications in the 2000s versus 1950s to 1970s, it appears that the number of
41
42 keywords for articles is also slowly growing over time. This trend may indicate the evolution in
43
44 computing possibilities to be able to publish large amounts of data, but also may indicate a desire
45
46 to combine datasets so that even more stimuli may be considered at once for modeling or
47
48 experiment creation.
49
50
51
52

53 Next, tags with at least a sample of 30 publications were investigated individually for
54
55 trends across time (correlations presented in Table 3, all $ps = 0.29-0.40$). Individual graphs can
56
57
58
59
60

LINGUISTIC BIBLIOGRAPHY

15

1
2
3 be created by clicking on the Tags Per Year Graph link online, which show the total frequency of
4 the selected tag by year. While correlations were not significant, some small positive trends were
5 found, such as the increase in age of acquisition, familiarity, imagery, orthographic
6 neighborhood, syllables, and valence norms. Intriguingly, meaningfulness and association both
7 showed negative correlations, but these correlations can be understood as an artifact of the
8 publication of a book on association norms in the 1970s (Postman & Keppel, 1970), as well as a
9 recent drop off of in the small but steady use of meaningfulness. These small correlations may
10 partially be explained by the sheer number and variation of data available in the LAB portal, as
11 one would expect the number of frequency tags to increase with the recent SUBTLEX
12 publications. Indeed, if the frequency tags are plotted by year an increase across the last decade
13 (22 tags in 2013 alone) can be shown. Readers are encouraged to view the individual graphs for
14 tags to investigate the change of keyword frequency over time, including the rise and demise of
15 several research areas. For example, confusion matrices heyday appears to range from the early
16 70s to the mid 80s, while arousal norms do not make an appearance until the late 90s.

36 **Conclusion**

37
38 This article had two main purposes: 1) to present the LAB dataset and portal as an
39 annotated bibliography and searchable tool for researchers, and 2) to view trends in
40 psycholinguistic research with an eye toward big data. We believe the LAB website will be a
41 useful channel for all levels of researchers, from graduate students looking for experimental
42 stimuli to design their experiments, to the familiar investigator who wishes to dig deeper into the
43 diverse choices offered. Further, while the majority of publications occur in one particular
44 journal, the LAB allows someone to find articles they may have missed in other areas with the
45 advantage of being collected into one location. User-friendly search tools are provided to aide in
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

LINGUISTIC BIBLIOGRAPHY

16

1
2
3 searching for specific languages, stimuli, or keywords, as well as multiple outputs for easy
4
5 copying into Excel or SPSS. With the advent of DOI, links are provided to the original source for
6
7 a quick transition to the materials provided in that publication. While this article's statistics will
8
9 become dated with the updates to the LAB, dynamic tables and graphs are provided online to see
10
11 the current status of the field. Lastly, we encourage users to actively report errors and suggest
12
13 updates for the LAB dataset as a way to crowd source information that is surely missing,
14
15 especially in non-English languages.
16
17
18

19
20 In the introduction, we provided two examples of current megastudies (SUBTLEX and
21
22 the Lexicon projects), in addition to how researchers might collect big data through Mechanical
23
24 Turk or Twitter. This article stepped back from looking at individual, large studies or ways to
25
26 collect data to use the information provided by publications as a window into the fluctuations of
27
28 the field. Megastudies have become a prevalent topic, but data could have revealed that this
29
30 popularity was due to recent publication of a small subset of articles. Instead, analyses showed
31
32 that not only are the numbers of publications accumulating, but the sizes of datasets are also
33
34 growing in tandem. Megastudies specifically focus on large datasets, but big data can also be
35
36 indicated here by the divergence in languages available, number of places to publish such data,
37
38 and the increasing number of keywords for articles across years. Time will tell if these trends can
39
40 and will continue or if certain areas will see a confusion matrix type decline after many large
41
42 datasets are published. With the move of traditional lab experiments to smartphone and tablet
43
44 technology (Dufua et al., 2011), it seems likely that researchers in psycholinguistics will
45
46 continue to find new and creative ways to modernize the field.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, 17, 814-823. doi: 10.1111/j.1467-9280.2006.01787.x.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database [CD-ROM]. Philadelphia: Linguistic Data Consortium. University of Pennsylvania.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445-459, doi: 10.3758/BF03193014.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, 133, 283-316. doi: 10.1037/0096-3445.133.2.283.

Barca, L., Burani, C., & Arduino, L.S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers*, 34, 424-434, doi: 10.3758/BF03195471.

Boudelaa, S. & Marslen-Wilson, W. D. (2010). Aralex: A lexical database for Modern Standard Arabic. *Behavior Research Methods*, 42, 481-487, doi: 10.3758/BRM.42.2.481.

Bradshaw, J. L. (1984). A guide to norms, ratings, and lists. *Memory & Cognition*, 12, 202-206. doi: 10.3758/BF03198435.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE*, 5, e10773. doi: 10.1371/journal.pone.0010773.

LINGUISTIC BIBLIOGRAPHY

18

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- 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, doi: 10.3758/BRM.41.4.977.
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bolte, J., & Bohl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58, 412-424, doi: 10.1027/1618-3169/a000123.
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (online first, 2013). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 1-8, doi: 10.3758/s13428-013-0403-5.
- Buchanan, E. M., Holmes, J. L., Teasley, M. L., & Hutchison, K. A. (2013). English semantic word-pair norms and a searchable Web portal for experimental stimulus creation. *Behavior Research Methods*, 45, 746-757, doi: 10.3758/s13428-012-0284-z.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data?. *Perspectives on Psychological Science*, 6, 3-5. doi: 10.1177/1745691610393980.
- Burgess, C. & Livesay, K. (1998). The effect of corpus size in predicting reaction time in a basic word recognition task: Moving on from Kučera and Francis. *Behavior Research Methods, Instruments, & Computers*, 30, 272-277. doi: 10.3758/BF03200655.
- Cai, Q. & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, 5, e10729. doi: 10.1371/journal.pone.0010729.
- Cohen-Shikora, E. R., Balota, D. A., Kapuria, A., & Yap, M.J. (2013). The past tense inflection

LINGUISTIC BIBLIOGRAPHY

19

project (PTIP): speeded past tense inflections, imageability ratings, and past tense consistency measures for 2,200 verbs. *Behavior Research Methods*, 45, 151-159. doi: 10.3758/s13428-012-0240-y.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, 132, 163-201. doi: 10.1037/0096-3445.132.2.163.

Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science*, 23, 371-414. doi: 10.1016/S0364-0213(99)00005-1.

Cuetos, F., Glez-Nosti, M., Barbon, A., & Brysbaert, M. (2011). SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicologica*, 32, 133-143.

De Deyne, S. L. S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, 45, 480-498, doi: 10.3758/s13428-012-0260-7.

Dimitropoulou, M., Dunabeitia, J.A., Aviles, A., Corral, J., & Carreiras, M. (2010). Subtitle-based word frequencies as the best estimate of reading behavior: the case of Greek. *Frontiers in Psychology*, 1, 1-12, doi: 10.3389/fpsyg.2010.00218.

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: hedonometrics and twitter. *PLoS ONE*, 6, e26752. doi: 10.1371/journal.pone.0026752.

Dufau, S., Duñabeitia, J. A., Moret-Tatay, C., McGonigal, A., Peeters, D., Alario, F. X., ... & Grainger, J. (2011). Smart phone, smart science: how the use of smartphones can

LINGUISTIC BIBLIOGRAPHY

20

- revolutionize research in cognitive science. *PloS ONE*, 6, e24974. doi:
10.1371/journal.pone.0024974.
- Guasch, M., Boada, R., Ferre, P., & Sanchwz-Casas, R. (2013). NIM: A Web-based Swiss army
knife to select stimuli for psycholinguistic studies. *Behavior Research Methods*, 45, 765-
771. doi: 10.3758/s13428-012-0296-8.
- Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C., ... &
Buchanan, E.M. (2013). The semantic priming project. *Behavior Research Methods*, 45,
1099-1114. doi: 10.3758/s13428-012-0304-z.
- Jasmin, K. & Casasanto, D. (2012). The QWERTY Effect: How typing shapes the meanings of
words. *Psychonomic Bulletin & Review*, 19, 499-504. doi: 10.3758/s13423-012-0229-7.
- Kent, G. H. & Rosanoff, A. J. (1910). A study of association in insanity. *The American Journal
of Psychiatry*, 67, 317-390. doi: 10.1037/13767-000.
- Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word
frequency based on film subtitles. *Behavior Research Methods*, 42, 643-650. doi:
10.3758/BRM.42.3.643.
- Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual
word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic
words and nonwords. *Frontiers in Psychology*, 1, 174. doi: 10.3389/fpsyg.2010.00174.
- Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project:
Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior
Research Methods*, 44, 287-304. doi: 10.3758/s13428-011-0118-4.
- Kloumann, I. M., Danforth, C. M., Harris, K. D., Bliss, C. A., & Dodds, P. S. (2012). Positivity
of the English language. *PLoS ONE*, 7, e29484. doi: 10.1371/journal.pone.0029484.

LINGUISTIC BIBLIOGRAPHY

21

Kucera, H. & Francis, W. N. (1967). Computational analysis of present-day American English. Providence, RI: Brown University Press.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44, 978-990. doi: 10.3758/s13428-012-0210-4.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240. doi: 10.1037//0033-295X.104.2.211.

Lete, B., Sprenger-Charolles, L., & Cole, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers*, 36, 156-166. doi: 10.3758/BF03195560.

Maki, W. S., McKinley, L. N., & Thompson, A. G. (2004). Semantic distance norms computed from an electronic dictionary (WordNet). *Behavior Research Methods, Instruments, & Computers*, 36, 421-431. doi: 10.3758/BF03195590.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44, 1-23. doi: 10.3758/s13428-011-0124-6.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, 126, 99-130. doi: 10.1037//0096-3445.126.2.99.

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, 7, 141-144. doi: 10.1016/S1364-6613(03)00029-9.

Moss, H. E., Tyler, L. K., & Devlin, J. T. (2002). The emergence of category-specific deficits in a distributed semantic system. In E. Forde & G Humphreys (Eds.) *Category Specificity in*

LINGUISTIC BIBLIOGRAPHY

22

- Brain and Mind*, 115-147. New York: Psychology Press.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36, 402-407. doi: 10.3758/BF03195588.
- New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, 28, 661-677. doi: 10.1017/S014271640707035X.
- Pexman, P. M., Holyk, G. G., & Monfils, M. H. (2003). Number-of-features effects and semantic processing. *Memory & Cognition*, 31, 842-855. doi: 10.3758/BF03196439.
- Postman, L. & Keppel, G. (1970). *Norms of word association*. New York: Academic Press.
- Proctor, R. W., & Vu, K. P. L. (1999). Index of norms and ratings published in the Psychonomic Society journals. *Behavior Research Methods, Instruments, & Computers*, 31, 659-667. doi: 10.3758/BF03200742.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14, 191-201. doi: 10.3758/BF03197692.
- Rogers, T. T., & McClelland, J. L. (2004). Semantic cognition: A parallel distributed processing approach. Boston: The MIT press.
- Snodgrass, J. G. & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 174-215. doi: 10.1037/0278-7393.6.2.174.
- Soares, A. P., Medeiros, J. C., Simoes, A., Machado, J., Costa, A., Iriarte, A., ... & Comesana M. (online first, 2013). ESCOLEX: A grade-level lexical database from European

LINGUISTIC BIBLIOGRAPHY

23

Portuguese elementary to middle school textbooks. *Behavior Research Methods*, 1-4. doi: 10.3758/s13428-013-0350-1.

Sze, W., Rickard Liow, S. J., & Yap, M. J. (*online first*, 2013). The Chinese Lexicon Project: A repository of lexical decision behavioral responses for 2,500 Chinese characters. *Behavior Research Methods*, 1-11. doi: 10.3758/s13428-013-0355-9.

van Heuven, W. J. B., Mander, P., Keuleers, E., & Brysbaert, M. (*online first*, 2013). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 1-36. doi: 10.1080/17470218.2013.850521.

Vaughan, J. (2004). Editorial: A web-based archive of norms, stimuli, and data. *Behavior Research Methods, Instruments, & Computers*, 36, 363-370. doi: 10.3758/BF03195583.

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, 48, 422-488. doi: 10.1016/j.cogpsych.2003.09.001.

Vinson, D. P., Vigliocco, G., Cappa, S., & Siri, S. (2003). The breakdown of semantic knowledge: Insights from a statistical model of meaning representation. *Brain and Language*, 86, 347-365. doi: 10.1016/S0093-934X(03)00144-5.

Vo, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, 41, 534-538. doi: 10.3758/BRM.41.2.534.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and

LINGUISTIC BIBLIOGRAPHY

24

- dominance for 13,915 English lemmas. *Behavior Research Methods*, 45, 1191-1207. doi: 10.3758/s13428-012-0314-x.
- Yap, M. J., Liow, S. J. R., Jalil, S. B., & Faizal, S. S. B. (2010). The Malay Lexicon Project: A database of lexical statistics for 9,592 words. *Behavior Research Methods*, 42, 992-1003. doi: 10.3758/BRM.42.4.992.
- Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin & Review*, 18, 742-750. doi: 10.3758/s13423-011-0092-y.
- Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language*, 47, 1-29. doi: 10.1006/jmla.2001.2834.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Running head: LINGUISTIC BIBLIOGRAPHY

25

Table 1. *Frequency Chart of Stimuli Types*

Stimuli	<i>N</i>	%	<i>M (SD)</i>	<i>Min</i>	<i>Max</i>	<i>Description</i>
Anagrams	6	0.95	229.00 (210.72)	80	378	Words whose letters can be rearranged into other real words.
Categories	27	4.27	37.58 (30.38)	4	100	Lists of words that are associated with particular category names, such as <i>animals</i> .
Characters	11	1.74	4278284.73 (14102958.94)	48	46800000	Characters are non-Roman letters, usually Japanese or Chinese logographs.
Cloze/Sentences	21	3.32	344.00 (438.04)	16	1998	Sentence norms are complete or partial sentences in structure. Cloze norms are sentences with a missing word, and probabilities of different words to complete that blank are provided.
Color drawings	6	0.95	325.67 (191.75)	200	714	Line drawings or similar non-picture images that are colored.
Homo/Heterographs	10	1.58	161.44 (175.01)	20	566	Homographs are two words with the same spelling, often with different pronunciations, and different meanings (<i>bow</i>), while heterographs have different spellings and meanings (<i>to</i> , <i>two</i>).
Homo/Heteronyms	4	0.63	265.67 (241.37)	114	544	Homonyms have the same spelling and pronunciation, but have different meanings (<i>bank</i>), while heteronyms have the same spelling with different pronunciations and meanings (<i>desert</i>).

LINGUISTIC BIBLIOGRAPHY

26

Homo/Heterophones	3	0.47	148.00 (93.66)	40	207	Homophones have the same pronunciation but different meanings (<i>rose</i>), while heterophones are often considered heteronyms with the same spelling but different pronunciations and meanings.
Letters	57	9.00	765.75 (1703.28)	19	8836	Alphabetic written elements.
Line drawings	34	5.37	237.79 (129.70)	22	520	A non-picture image that is not colored.
Names	6	0.95	3103.40 (4385.68)	126	10000	Words that are traditionally considered first or last names, such as <i>Bob</i> and <i>Smith</i> .
Other	34	5.37	311.00 (307.85)	1	811	This category was used for stimuli that did not fit into others, programs or experimental stimuli creation tools, and each is described in the online table with that particular row in stimuli notes.
Phonemes	3	0.47	10000.00 (0.00)	10000	10000	A basic unit of sound in a language wherein changes bring changes to a word's meaning.
Pictures	36	5.69	332.47 (279.36)	2	1095	Photographs or other complex images.
Proverbs	1	0.16	500.00 (0.00)	500	500	A short saying that states a general truth or piece of advice.
Pseudowords	8	1.26	18795.83 (16999.47)	2500	40481	Non-real words that are often created by changing one letter of a real word to create a linguistically valid consonant-vowel combination (<i>wug</i>).
Sounds	8	1.26	117.00 (48.51)	22	167	Clips of noises, speech, or songs.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

LINGUISTIC BIBLIOGRAPHY

Syllables	9	1.42	18129.00 (49247.04)	33	140000	A unit of pronunciation with at least one vowel, diphthong, or consonant.
Symbols/Icons	4	0.63	211.67 (132.48)	68	329	Non-word characters.
Word Pairs	25	3.95	9196.05 (22145.54)	40	72186	Words that were specifically paired for study in the article, such as studies on priming or word association.
Words	320	50.55	7674664.48 (49139640.06)	10	593000000	A distinct meaningful element of speech or writing.

LINGUISTIC BIBLIOGRAPHY

28

Table 2. *Frequency Chart of Language Types*

Language	<i>N</i>	<i>Percentage</i>
British English	20	3.57
Chinese	15	2.67
Dutch	10	1.78
English	359	63.99
French	32	5.70
German	20	3.57
Greek	5	0.89
Italian	11	1.96
Japanese	10	1.78
Multiple	23	4.10
Other	9	1.60
Portuguese	10	1.78
Spanish	37	6.60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

LINGUISTIC BIBLIOGRAPHY

Table 3. *Frequency and correlation of tags*

Stimuli	<i>N</i>	%	<i>r</i> With Year	Description
Age of Acquisition	75	13.37	0.21	Estimated age of first learning for a concept.
Ambiguity/Word Meaning	30	5.35	-0.01	Estimates of ambiguity for a concept or information about different word meanings (i.e. number of meanings).
Arousal	29	5.17	NA	Estimates of strength of response to a concept.
Association	81	14.44	-0.30	Estimates of the relationship between concepts that are used together, often in free association format (i.e. the first word that pops to mind after a cue word has been listed).
Category	40	7.13	-0.02	Information related to lists of words all related to one cue word, such as <i>bird</i> , <i>plant</i> , <i>animal</i> .
Cloze Probabilities	5	0.89	NA	The probability of an individual word completing a specific spot in a sentence.
Complexity	14	2.50	NA	Estimates of the intricacy or complicatedness of a concept.
Concreteness	55	9.8	0.03	Estimates of the non-abstractness of a concept, sometimes described as how touchable a concept is.
Confusion Matrices	18	3.21	NA	Probabilities of distinctiveness between pairs of concepts, or the likelihood those two concepts will be confused with each other (usually letters).
Danger	1	0.18	NA	Estimates of how dangerous, scary a concept seemed.
Distinctiveness	7	1.25	NA	Estimates of how different, unique a concept seemed.
Dominance	26	4.63	NA	Estimates of how important or powerful a concept seemed.
Ease of Learning	5	0.89	NA	Estimates of how difficult a concept was to remember.
Familiarity	97	17.29	0.19	Estimates of how well known a concept seemed.
Frequency	190	33.87	0.09	The rates of occurrence for concepts.
Grapheme-Phoneme Correspondence	17	3.03	NA	The relationship between written and spoken symbols.
Identification	11	1.96	NA	Estimates of the likelihood of remembering concepts.
Identification - Lexical	7	1.25	NA	Participant response times to real and pseudowords in a

LINGUISTIC BIBLIOGRAPHY

30

Decision				yes/no decision task.
Identification - Naming	32	5.70	0.14	Participant response times to reading aloud real and pseudowords.
Image Agreement	14	2.50	NA	Estimates of the similarity between images.
Image Variability	5	0.89	NA	Estimates of the complexity of images.
Imagery	14	5.35	0.19	Estimates of how easy or difficult a mental image is to create (situation based), older term related to imageability.
Imageability	49	8.73	-0.09	Estimates of how easy or difficulty a concept is to imagine.
Intensity	2	0.36	NA	Estimates of the strength of a concept or how intense the concept seems.
Letters	46	8.20	0.11	Alphabetic characters.
Meaningfulness	36	6.42	-0.19	Estimates of how meaningful or significant a concept seems.
Modality	4	0.71	NA	The different ways or modes that something can be experienced (i.e. perceptual, tactile).
Morphology	8	1.43	NA	The structure of a concept's morphemes.
Name Agreement	28	4.99	NA	Estimates of the similarity between concepts.
Orthographic Neighborhood	40	7.13	0.17	The number of concepts that differ by one letter.
Part of Speech	41	7.31	0.11	Type of speech for a concept (i.e. verbs, nouns).
Phonemes	35	6.24	NA	A basic unit of sound in a language wherein changes bring changes to a word's meaning.
Phonological Neighborhood	26	4.63	NA	The number of concepts that differ by one phoneme.
Priming	4	0.71	NA	The advantage in processing a concept receives when shown together with a related concept.
Pronunciation	16	2.85	NA	The way in which a concept is spoken.
Response Times	52	9.27	0.09	The amount of time required to respond to a concept.
Recall	17	0.36	NA	The likelihood of remembering a concept from memory.
Recognition	12	2.14	NA	The likelihood of recognizing (remembering) a concept from memory.

LINGUISTIC BIBLIOGRAPHY

Rime	5	0.89	NA	Concepts that have corresponding sounds.
Segmentation	1	0.18	NA	The separation of concepts into parts.
Semantics	74	13.19	0.09	Estimates of relationships between word pairs based on definition and meaning.
Sensory/Motor	23	4.10	NA	Studies that used different sensory mechanisms (tactile or visually presented) or had ratings of actions/animation of concepts.
Sentence Completion	6	1.07	NA	Estimates of common conceptions that would complete a blank in sentence.
Similarity	13	2.32	NA	Estimations of the likeness or sameness between two concepts.
Syllables	35	6.24	0.17	A unit of pronunciation with at least one vowel, diphthong, or consonant.
Typicality	20	3.57	NA	Estimates of how common or normal a feature is for a concept.
Valence/Emotion	59	10.52	0.12	Estimates of the strength of emotions or feelings for concepts.
Visual Complexity	24	4.28	NA	Estimates of the intricacy or complicatedness of a picture.
Word Completion	9	1.60	NA	Estimates of likelihoods of letter combinations that would complete a word.

Note. Only tags with 30 or more occurrences have correlation calculations.

LINGUISTIC BIBLIOGRAPHY

32

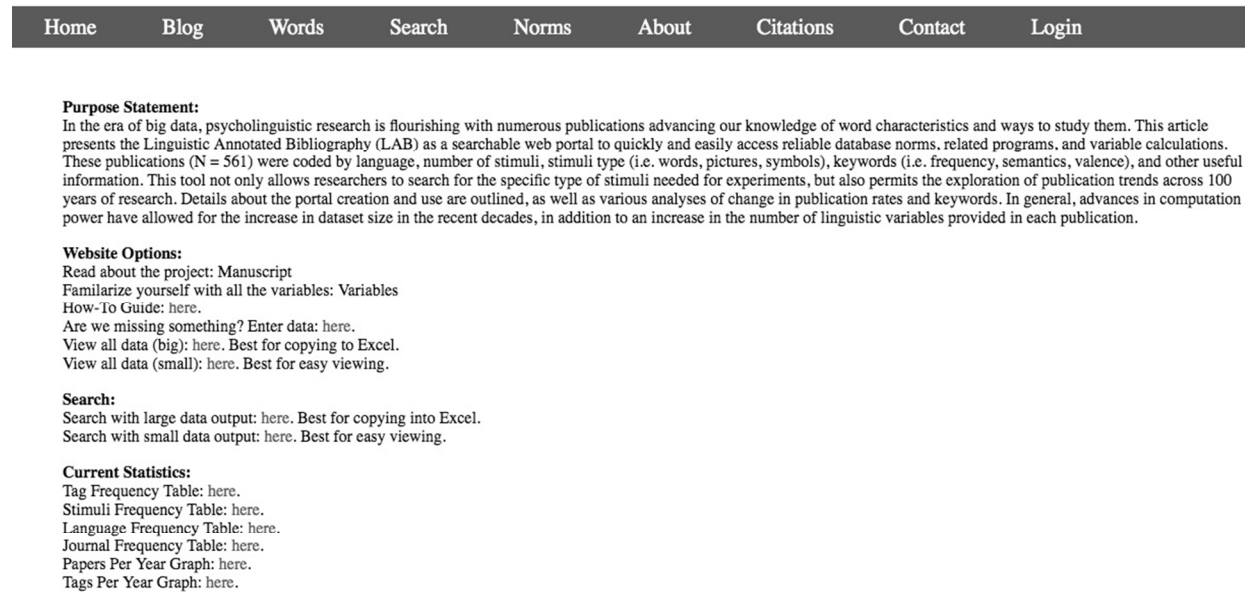


Figure 1. The layout of the LAB web portal, which can be found by clicking on the Norms link in the menu bar.

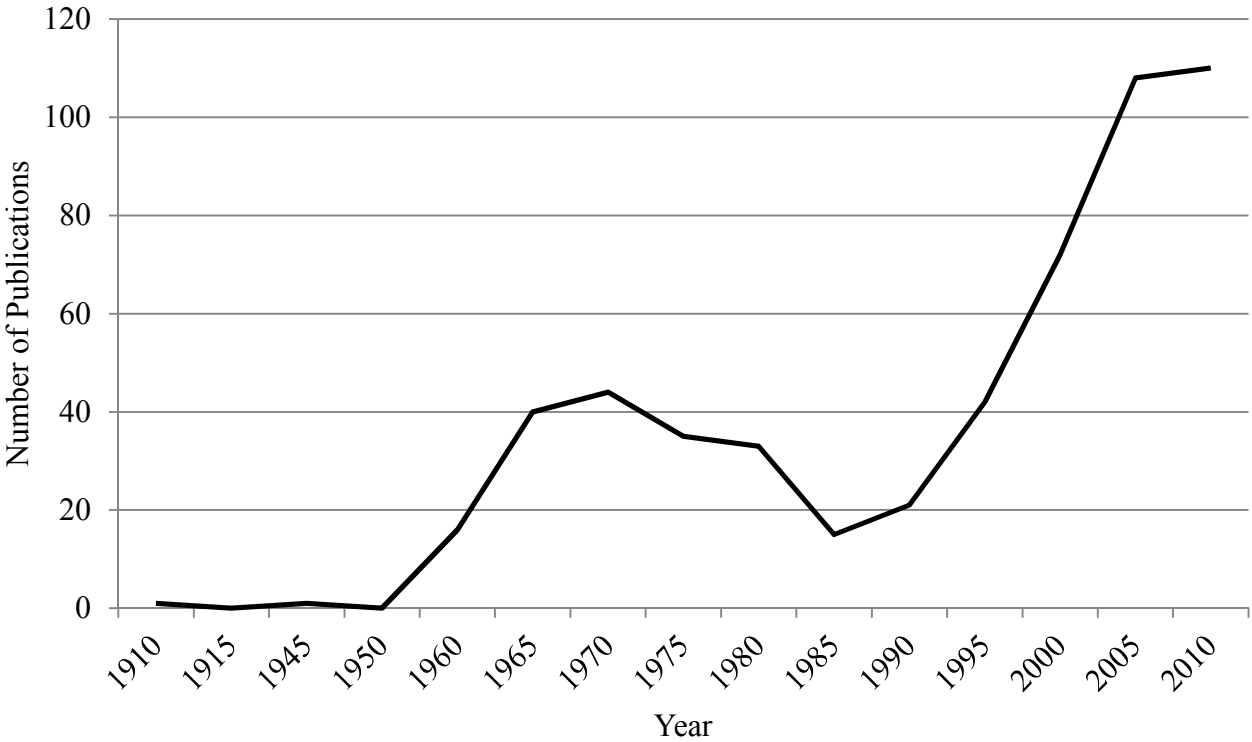
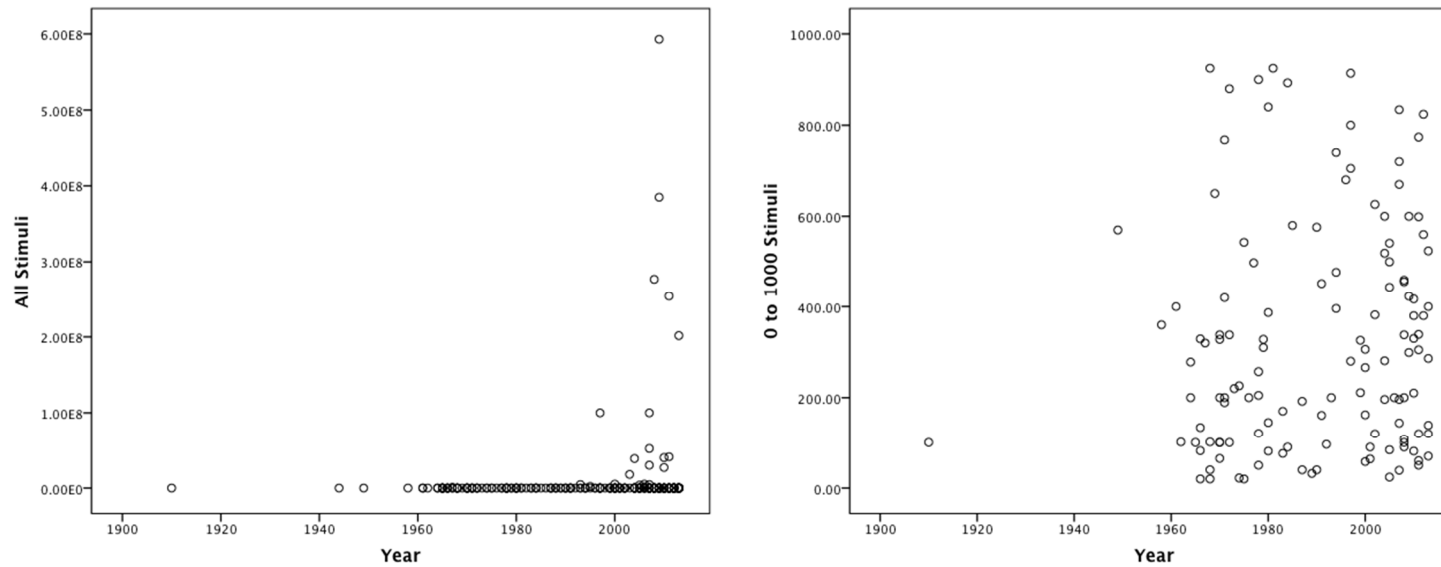


Figure 2. Number of publications plotted in half decade intervals.

LINGUISTIC BIBLIOGRAPHY

34



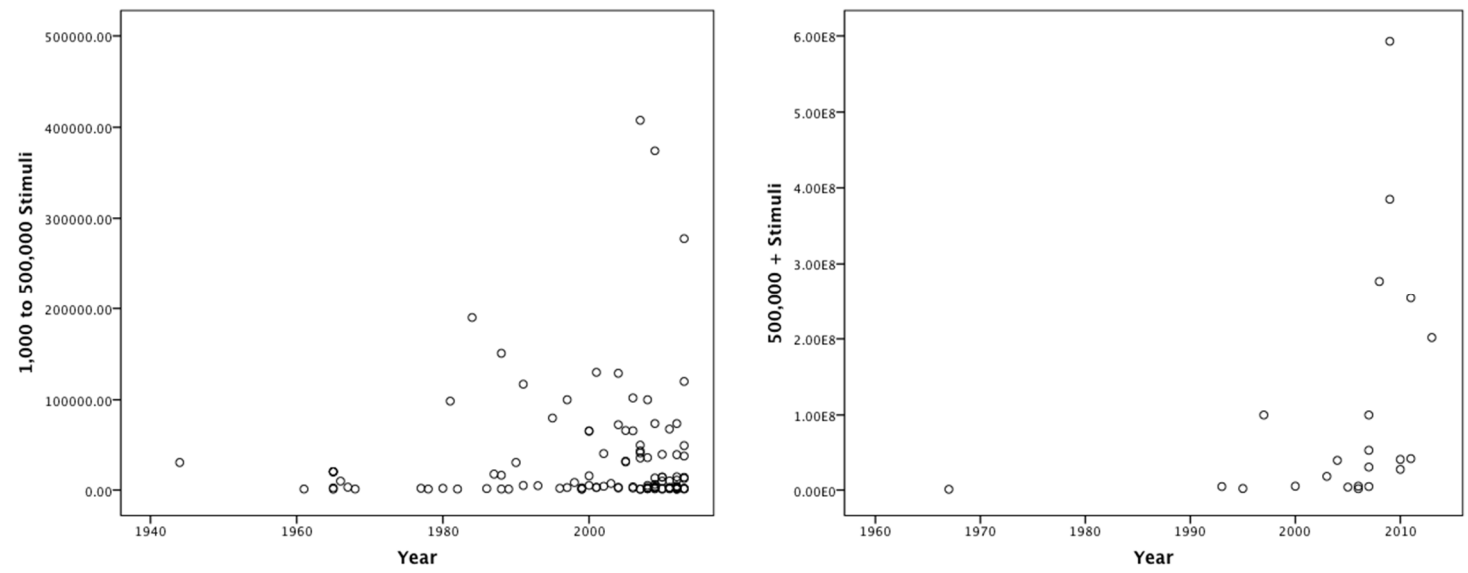


Figure 3. Scatterplots of number of stimuli by year. The top left quadrant shows all years and stimuli, the top right quadrant shows stimuli numbers in the 0 to 1000 range, and the bottom two quadrants show 1,000 – 500,000, and 500,000+ numbers of stimuli across years.

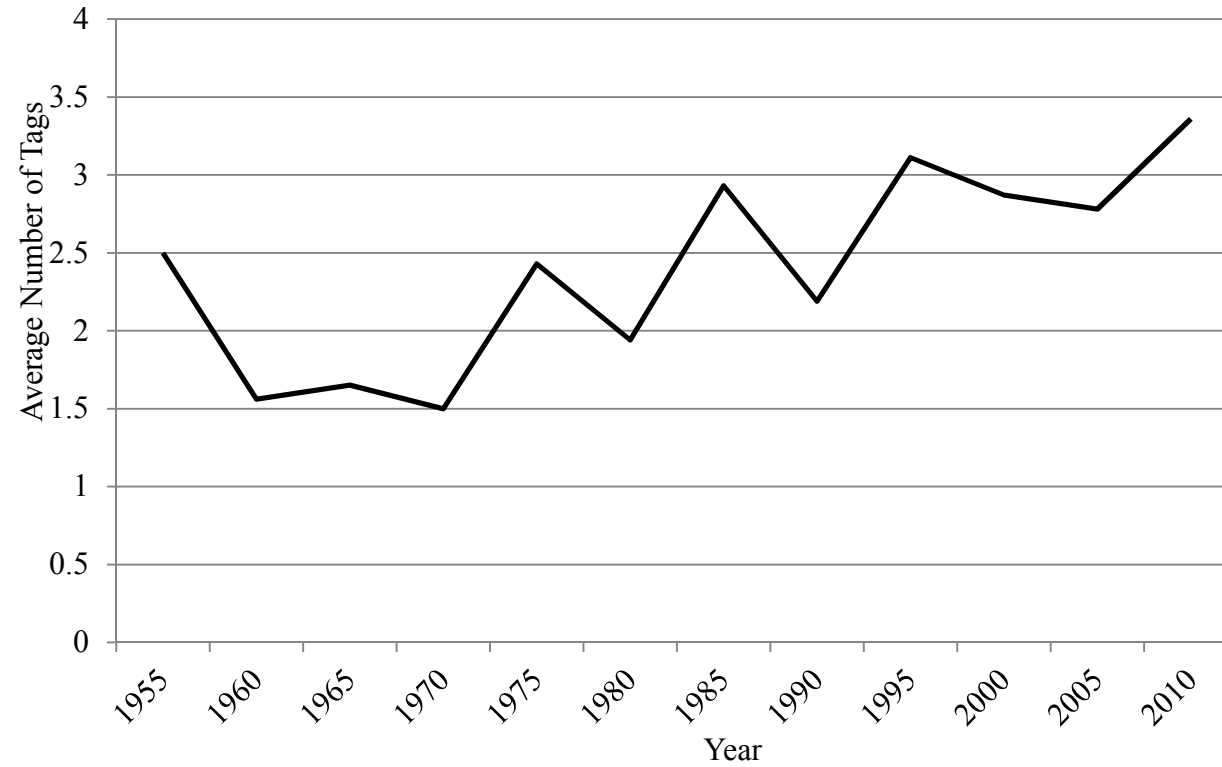


Figure 4. Average number of tags per publication plotted in half decade intervals.