English Semantic Feature Production Norms: An Extended Database of 4,437 Concepts

Erin M. Buchanan[1], K. D. Valentine[2], & Nicholas P. Maxwell[1]

[1] Missouri State University

[2] University of Missouri

Author Note

Erin M. Buchanan is an Associate Professor of Quantitative Psychology at Missouri State University. K. D. Valentine is a Ph.D. candidate at the University of Missouri. Nicholas P. Maxwell is a Masters' candidate at Missouri State University.

We would like to thank Keith Hutchison and David Balota for their contributions to this project, including the funds to secure Mechanical Turk participants.

Correspondence concerning this article should be addressed to Erin M. Buchanan, 901 S. National Ave, Springfield, MO 65897. E-mail: erinbuchanan@missouristate.edu

Abstract

The largest limiting factor in understanding memory and language networks is often the availability of normed stimuli to use and explore in experimental studies. In this study, we expand on three previous semantic feature overlap norms to over 4,000 cue stimuli ranging from nouns, verbs, adjectives, and other parts of speech. Participants in the norming study were asked to provide feature components of each cue stimuli, which were combined with the previous research using semantic feature production procedures. In addition to expanding previous research, this project explores different semantic overlap measurements by coding each word feature listed by root and affixes to determine different strengths of feature overlap. All information is provided in a searchable database for easy access and utilization for future researchers when designing experiments. The final database of cue-target pairs was paired with the Semantic Priming Project to examine the predictive ability of semanticity on semantic priming in tandem with other psycholinguistic variables, such as association, thematics, and frequency. Target concept frequency was the largest predictor of semantic priming, follow by thematics and association. Root word cosine was predictive of semantic priming, even after adjusting for the previously mentioned psycholinguistic variables.

Keywords: semantics, word norms, database, psycholinguistics

English Semantic Feature Production Norms: An Extended Database of 4,437 Concepts

Semantic representations are the focus of a large area of research and are a thing.

What are semantic feature production norms

Why are they important!

Previous work

Method

Participants

Participants in the newly collected stimuli were gathered from Amazon's Mechanical Turk, which is a large, diverse participant pool wherein users can complete surveys for small sums of money (Buhrmester, Kwang, & Gosling, 2011). Answers can be screened for errors, and incorrect or incomplete surveys can be rejected or discarded without payment. Each participant was paid five cents for a survey, and they could complete multiple Human Intelligence Tasks or HITS. Each HIT included five concepts, and HITS would remain active until $n = 30$ valid survey answers were collected. HITS were usually rejected if they included copied definitions from Wikipedia, "I don't know", or writing a paragraph about the concept. These answers were discarded, as described below. Table 1 includes the sample sizes from the new study (Mechanical Turk 2), as well as the sample sizes from the previous study, as described in Buchanan, Holmes, Teasley, and Hutchison (2013).

Materials

The purpose of this second norming set was to expand the Buchanan et al. (2013) norms to include all concepts from the Semantic Priming Project (**???**). Therefore, these concepts were the target of the project. The original concept set was selected primarily from the Nelson, McEvoy, and Schreiber (2004) database, with small overlaps in the McRae, Cree, Seidenberg, and McNorgan (2005) and Vinson and Vigliocco (2008) database sets for convergent validity. In the Semantic Priming Project, cue-target pairs were shown to participants to examine naming and lexical decision time priming across related and unrelated pairs. The related pairs included first associate (most common response to a cue) and other associates (second or greater common responses to cues) as their target words. The original publication of concepts included the cue words from the Semantic Priming Project, while this project expanded to include missed cue words and all target words. The addition of these concepts allowed for complete overlap between the Semantic Priming Project and the feature production norms.

Concepts were labeled by part of speech using the English Lexicon Project (Balota et al., 2007), the free association norms, and Google's define search when necessary. When labelling these words, we used the most common part of speech to categorize concepts. This choice was predominately for simplicity of categorization, however, the participants were shown concepts without the suggestion of which sense to use for the word. Therefore, multiple senses are embedded into the feature production norms, while the database is labeled with single parts of speech. The other parts of speech can be found in the English Lexicon Project or multiple other databases. This dataset was combined with (**???**) and Vinson and Vigliocco (2008) feature production norms, which was a combined total of 4437 concepts. 70.4% of concepts were nouns, 14.9% adjectives, 12.4% verbs, and 2.3% were other forms of speech, such as adverbs and conjunctions.

Procedure

Each HIT was kept to five concepts, and usual survey response times were five to seven minutes. Each HIT was open until thirty participants had successfully completed the HIT and were paid the five cents for the HIT. The survey instructions were copied from McRae et al. (2005)'s Appendix B, which were also used in the previous publication of these norms. Because the McRae et al. (2005) data was collected on paper, we modified these instructions slightly. The original lines to write in responses were changed to an online textbox response window. The detailed instructions additionally no longer contained information about how a participant should only consider the noun of the target concept, as the words in our study included multiple forms of speech and senses. Participants were encouraged to list the properties or features of each concept in the following areas: physical (looks, sounds, and feels), functional (uses), and categorical (belongings). The same examples used previously (*duck, cucumber, stove*) were included to aid in task understanding and completion. Participants signed up for the HITS through Amazon's Mechanical Turk website and completed the study within the Mechanical Turk framework. Approved HITs were compensated through the Mechanical Turk system. All answers were then combined into a larger dataset.

Data Processing

The entire dataset, at each processing stage described here, can be found at: https://osf.io/cjyzw/. On our OSF page, we have included a detailed processing guide on how concepts were (re)examined for this publication. This paper was written with *R* markdown and *papaja* (Aust & Barth, 2018). The markdown document allows an interested reader to view the scripts that created the article in line with the written text. However, the processing of the text documents was performed on the raw files, and therefore, we have included the processing guide for transparency of each stage.

First, each concept was separated into an individual text file that is included as the "raw" data online. Each of these files was then spell checked and corrected when the participant answer was obviously a typo. As noted earlier, participants often tried to cut and paste Wikipedia or other online dictionary sources into the their answers to complete surveys quickly with minimal effort. These entries were easily found by the formatting of the webpage that was included in their answer. These answers were then discarded from the concept individual text files. Next, each concept was processed for feature frequency. In this stage, the raw frequency counts of each cue-feature combination were calculated and put together into one large file. Cue-cue combinations were discarded, as participants might write "a zebra is a horse" when asked to define *zebra*. English stop words such as *the, an, of* were then discarded, as well as terms that were often used as part of a definition (*like, means, describes*).

To create the final root cosine values, we then created a "translated" column for each feature listed. This column indicated the root word for each feature, along with the affixes that were used in the original feature. For example, the original feature *cats* would be translated to *cat* and *s*, wherein *cat* would be the translated feature and the *s* would be the affix code. Multiple affix codes were often needed for features, as *beautifully* would have been translated to *beauty*, *ful*, and *ly*. Often, the noun version of the feature would be used for the translation or the most common part of speech for each feature would be recorded. The sample size for the cue was added to this dataset, as the sample sizes varied across experiment time, as shown in Table 1. Therefore, instead of using raw feature frequency, we normalized each count into a percent of participants that included that feature with each cue.

At this stage, the data was reduced to cue-feature combinations that were listed by at least 16% of participants (matching McRae et al. (2005)'s procedure) or were in the top five features listed for that cue. This calculation was performed on the translated normalized feature percent. For example, *beauty* may have been listed as *beauty, beautiful, beautifully, beautifulness*, and this feature would have been listed three times in the dataset for the original cue. The *frequency_feature* column indicates the frequency of the original, unedited feature, while the *frequency_translated* includes all combinations of *beauty* into one overall feature. Because non-nouns can be more difficult to create a feature list for, we included the top five descriptors in addition to the 16% listed criteria, to ensure that each concept included at least five features. Table 2 indicates the average number of cue-feature pairs found for each data collection site/time point and part of speech for the cue word.

The parts of speech for the cue, original feature, and translated feature were merged with this file as described above. Table 3 depicts the pattern of feature responses for cue-feature part of speech combinations. This table includes the percent of features listed for each cue-feature part of speech combination (i.e., what is the percent of time that both the cue and feature were both adjectives) for the original feature (raw) and translated feature (root). Next, the normalized frequency percent average was calculated along with their standard deviations. These columns indicate the frequency percent that a cue-feature part of speech combination was listed across participants (i.e., what is the average percent of participants that listed an adjective feature for an adjective cue). These two types of calculation describe the likelihood of seeing part of speech combinations across the concepts, along with the likelihood of those cue-feature part of speech combinations across participants.

The top cue-feature combinations from Buchanan et al. (2013) and this new data collection were then combined with the cue-feature combinations from McRae et al. (2005) and Vinson and Vigliocco (2008). We did not reduce their cue-feature combinations, but instead included them with the cue-feature listed in their supplemental files with the feature in the raw feature column. If features could be translated into root words with affixes, the same procedure as described above was applied. The final file then included the original dataset, cue, feature, translated feature, frequency of the original feature, frequency of the translated feature, sample size, normalized frequencies for the original and translated feature. This file includes 69284 cue-feature combinations, with 48925 from our dataset, and 24449 of which are unique cue-translated feature combinations. Statisics in Tables 2 and 3 only include information from the reprocessed Buchanan et al. (2013) norms and the new cues collected for this project. The final data processing step was to code affixes found on the original features. A complete affix list translation file can be found online in our OSF files. Table 4 displays the list of affix tags, common examples for each type of affix, and the percent of affixes that fell into each cateory. The percent values are calculated on the overall affix list, as feature words could have up to three different affixes. Generally, affixes were tagged in a one-to-one match, however, special care was taken with numbers and verb tenses. Features like cat*s* would be coded as a number affix, while features like walk*s* would be coded as a third person verb. In the final words file found online, we additionally added forward strength (FSG) and backward strength (BSG) for investigation into association overlap. The last few columns indicate the word list a concept was originally normed in to allow for matching to the original raw files on the OSF page, along with the code for each school and time point of collection.

This procedure is a slight departure from our previous work, as we previously argued to keep some morphologically similar features separate if they denoted different concepts. For example, *act* and *actor* were separated because each feature explained a separate component of the cue word (i.e., noun and gender). The original processing in Buchanan et al. (2013) combined features that overlapped in cue sets by 80%. In this reprocessing and update, we translated all words to a root form, and coded these translations, thus, allowing for the exploration of the affect of affixes on semantic feature overlap. Both forms of the feature are provided for flexibility in calculating overlap by using the original feature (raw), the translated feature (root), and the affix overlap by code (affix). Cosine values were calculated for each of these feature sets by using the following formula:

$$\frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}$$

This formula is similar to a dot-product correlation, where A_i and B_i indicate the overlapping feature normalized frequency between cue A and cue B. The i subscript denotes the current cue, and when features match, the frequencies are multiplied together and summed across all matches (Σ). For the denominator, the feature normalized frequency is first squared and summed from i to n features for cue A and B. The square root of these summation values is then multiplied together. In essence, the numerator calculates the overlap of feature frequency for matching features, while the denominator accounts for the entire feature frequency set for each cue. Cosine values range from 0 (no overlapping features) to 1 (complete overlapping features). With nearly five thousand cue words, just under twenty-five million cue-cue cosine combinations can be calculated. In the datasets presented online, we only included cue-cue combinations with a feature overlap of at least two features, in order to reduce the large quantity of zero and very low cosine values. This procedure additionally allowed for online presentation of the data, as millions of lines was not feasible for our server. The complete feature list, along with our code to calculate cosine, can be used to obtain values not presented in our data if necessary.

Website

In addition to our OSF page, we present a revamped website for this data at http://www.wordnorms.com/. The single words page includes information about each of the cue words including cue set size, concreteness, word frequency from multiple sources, length, full part of speech, orthographic/phonographic neighborhood, and number of phonemes, syllables, and morphemes. These values were taken from Nelson et al. (2004), Balota et al. (2007), and (**???**). A definition of each of these variables is provided along with the minimum, maximum, mean, and standard deviation of numeric values. The table is programmed using Shiny apps (Chang, Cheng, Allaire, Xie, & McPherson, 2017). Shiny is an *R* package that allows the creation of dynamic graphical user interfaces for interactive web applications. The advantage to using Shiny applications is data manipulation and visualization with the additional bonus of up to date statistics for provided data (i.e., as typos are fixed or data is updated, the web app will display the most recent caculations). In addition to the variable table, users can search and save filtered output using our Shiny search app. With this app, you can filter for specific variable ranges and save the output in a csv or Excel file. The complete data is also provided for download.

On the word pairs page, all information about word-pair statistics can be found. A second variable table is provided with semantic and associative statistics. This dataset includes the cue and target words from this project (cue-cue combinations), the root, raw, and affix cosines described above, as well as the original Buchanan et al. (2013) cosines. Additional semantic information includes Latent Semantic Analysis (LSA; **???**) and JCN (**???**) values provided in the Maki, McKinley, and Thompson (2004) norms, along with FSG and BSG from the Nelson et al. (2004) norms for association. The descriptions, minimum, maximum, mean, and standard deviations of these values are provided in the app. Again, the search app includes all of these stimuli for cue-cue combinations with two or more features in common, where you can filter this data for experimental stimuli creation. The separation of single and word-pair data (as well as cosine calculation reduction to cues with two features in common) was practical, as the applications run slowly as a factor of the number of rows and columns. On each page, we link the data, applications, and source code so that others may use and manipulate our work depending on their data creation or visualization goals.

Results

An examination of the results of the cue-feature lists indicated that the new data collection is similar to the previous semantic feature production norms. As shown in Table 2, the new Mechanical Turk data showed roughly the same number of listed features for each cue concepts, usually between five to seven features. Table 3 portrayed that adjective cues generally included other adjectives or nouns as features, while noun cues were predominately described by other nouns. Verb cues included a large feature list of nouns, but then was equally split between adjectives, other verbs, and other categories. Lastly, the other cue types generally elicited nouns and verbs. Normalized percent frequencies were generally between seven and twenty percent of the participant sampling listing features when examining the raw words. These words included multiple forms, as the percent increased to around thirty percent when features were translated into their root words. Indeed, nearly half of the 48925 cue-feature pairs were repeated, as 24449 cue-feature pairs were unique when examining translated features.

36030 affix values were found, which was for 4407 cue concepts. 33052 first affixes were found, with 2832 second place affixes, and 146 third place affixes. Table 4 shows the distribution of these affix values. Generally, numbers were the largest category of affixes indicating that participants often indicated the quantity of the feature when describing the cue word. The second largest affix category was characteristics which often indicated the switch to or from a noun form of the feature word. Verb tenses (past tense, present participle, and third person) comprised a large set of affixes indicating the type of concept or when a concept might be doing an action for a cue. Persons and objects were also indicated about 7% of the time, while actions and processes of the cue were mentioned about 8% of the time.

Convergent Validity

To examine the validity of cosine values, we calculated the average cosine score between the new processing of the data for each of the three feature production norms used in this project. Overlapping cues in each of three database sets were found (n = 188), and the average cosine between their feature sets was examined. Buchanan et al. (2013) and the new dataset are listed the subscript B, while McRae et al. (2005) is M and V for Vinson and Vigliocco (2008). For root cosine values, we found high overlap between all three datasets: M_{BM} = .67 (SD = .14), M_{BV} = .66 (SD = .18), and M_{MV} = .72 (SD = .11). The raw cosine values also overlapped well, even though the McRae et al. (2005) and Vinson and Vigliocco (2008) datasets were already mostly preprocessed for word stems: M_{BM} = .55 (SD = .15), M_{BV} = .54 (SD = .20), and M_{MV} = .45 (SD = .19). Last, the affix cosines overlapped well between BM datasets, M_{BM} = .43 (SD = .29), but did not overlpa with the V datasets: M_{BV} = .04 (SD = .14), and M_{MV} = .09 (SD = .19). Last, the correlation between root, raw, affix, old cosine, LSA, and JCN were calculated to examine convergent validity. As shown in Table 5, the intercorrelations between the cosine measures are high, especially between our previous work and this dataset. JCN is backwards coded, as zero values indicate close semantic neighbors (low dictionary distance) and high values indicate low semantic relation. The small negative correlations replicate previous findings. LSA values showed small positive correlations with cosine values, indicating some overlap with thematic information and semantic feature overlap (Maki & Buchanan, 2008). These correlations are slightly different than our previous publication, likely because we restricted this cosine set to values with at least two features in common. The results are simiilar, where LSA and JCN correlations are lower than LSA-COS and JCN-COS, but these values indicate that themes and dictionary distance are similarly related to feature overlap.

Divergent Validity

Adjective Noun Verb Other 51.85612 36.47896 32.14514 44.44444 Adjective Noun Verb Other 0.1180191 0.1090159 0.1076604 0.1334710 Adjective Noun Verb Other 0.1495435 0.1379780 0.1344069 0.1827018 Adjective Noun Verb Other 0.010 0.010 0.010 0.013 Adjective Noun Verb Other 0.939 0.913 0.936 0.878

When collecting semantic feature production norms, there can be a concern that the information produced will simply mimic the free association norms, and thus, be a representation of association (context) rather than semanticity (meaning). Table 6 portrays the overlap with the Nelson et al. (2004) norms. The percent of time a cue-feature combination was present in the free association norms was calculated, along with the average FSG for those overlapping pairs. These values were calculated on the complete dataset with the McRae et al. (2005) and Vinson and Vigliocco (2008) norms, as we are presenting them as a combined dataset, on the translated cue-feature set only. The overall overlap between the database cue-feature sets and the free association cue-target sets was approximately 37%, ranging from 32% for verbs and nearly 52% for adjectives. Similiar to our previous results, the range of the FSG was large (.01 - .94), however, the average FSG was low for these overlapping pairs, $M = .11$ ($SD = .14$). These results indicated that while it will always be difficult to separate association and meaning, the dataset presented here represents a low association when examining overlapping values, and more than 60% of the data is completely separate from the free association norms. The limitation to this finding is the removal of idiosyncratic responses from the Nelson et al. (2004) norms, but even if these were to be included in some form, the average FSG would still be quite low when comparing cue-feature lists to cue-target lists.

Semantic Priming Prediction

A secondary goal of this

Discussion

With these values, we believe we have captured some of the nuisance that is often discarded in this type of research. (affixes)

References

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown.* Retrieved from https://github.com/crsh/papaja

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459. doi:10.3758/BF03193014

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*(3), 746–757. doi:10.3758/s13428-012-0284-z

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk. *Perspectives on Psychological Science*, *6*(1), 3–5. doi:10.1177/1745691610393980

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2017). *Shiny: Web application framework for r.* Retrieved from https://CRAN.R-project.org/package=shiny

Maki, W. S., & Buchanan, E. M. (2008). Latent structure in measures of associative, semantic, and thematic knowledge. *Psychonomic Bulletin & Review*, *15*(3), 598–603. doi:10.3758/PBR.15.3.598

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*(3), 421–431. doi:10.3758/BF03195590

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559. doi:10.3758/BF03192726

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*(3), 402–407. doi:10.3758/BF03195588

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, *40*(1), 183–190.

doi:10.3758/BRM.40.1.183

Table 1

Institution	Total Participants	Concepts	Mean *N*
University of Mississippi	749	658	67.8
Missouri State University	1420	720	71.4
Montana State University	127	120	63.5
Mechanical Turk 1	571	310	60
Mechanical Turk 2	198	1914	30

Note.

Table 2

Institution	Adjective	Noun	Verb	Other	Total
University of Mississippi	5.57 (1.53)	7.35 (4.05)	5.33 (.87)	6.01 (2.11)	6.71 (3.44)
Missouri State University	5.74 (1.56)	6.85 (2.82)	6.67 (2.08)	7.45 (5.35)	6.65 (2.92)
Montana State University	5.81 (1.74)	7.25 (3.35)	5.59 (1.13)	5.76 (1.74)	6.69 (2.93)
Mechanical Turk 1	6.27 (2.28)	7.74 (4.34)	5.77 (1.17)	5.57 (1.40)	7.14 (3.79)
Mechanical Turk 2	5.76 (1.36)	6.62 (1.85)	5.92 (1.38)	5.78 (1.17)	6.38 (1.75)
Total	5.78 (1.61)	6.94 (2.88)	5.67 (1.18)	5.84 (1.71)	6.57 (2.60)

Note.

Table 3

Cue Type	Feature Type	% Raw	% Root	*M* Freq. Raw	*M* Freq. Root
Adjective	Adjective	38.09	29.74	17.84 (16.47)	30.02 (18.83)
	Noun	40.02	46.74	13.14 (14.96)	29.71 (19.94)
	Verb	17.69	20.72	8.51 (9.78)	26.88 (17.27)
	Other	4.2	2.8	15.17 (15.64)	28.04 (15.54)
Noun	Adjective	16.56	12.07	15.55 (15.17)	31.20 (18.17)
	Noun	60.85	62.67	17.21 (17.01)	33.26 (20.05)
	Verb	20.8	23.68	8.88 (9.73)	31.01 (17.87)
	Other	1.79	1.58	17.06 (15.29)	28.87 (17.14)
Verb	Adjective	15.16	12.27	13.95 (13.98)	30.03 (18.28)
	Noun	42.92	44.35	14.59 (14.92)	29.59 (18.90)
	Verb	36.92	39.72	12.75 (14.85)	30.43 (19.54)
	Other	5	3.66	19.16 (15.95)	25.59 (19.54)
Other	Adjective	20.8	20.32	16.61 (17.37)	31.66 (19.51)
	Noun	42.74	39.03	16.77 (19.41)	37.28 (25.94)
	Verb	19.66	23.93	7.18 (7.57)	26.14 (19.38)
	Other	16.81	16.71	22.72 (16.69)	30.70 (18.48)
Total	Adjective	19.74	14.93	16.12 (15.57)	30.75 (18.37)
	Noun	55.41	57.81	16.55 (16.74)	32.58 (20.09)
	Verb	22.02	24.95	9.50 (10.91)	30.29 (18.24)
	Other	2.82	2.31	17.76 (15.83)	28.45 (16.83)

Note.

Table 4

Affix Tag	Example	Percent
Actions/Processes	ion, ment, ble, ate, ize	8.21
Characteristic	y, ous, nt, ful, ive, wise	22.72
Location	under, sub, mid, inter	0.44
Magnitude	er, est, over, super, extra	1.31
Not	less, dis, un, non, in , im, ab	2.76
Number	s, uni, bi, tri, semi	28.31
Opposites/Wrong	mis, anti, de	0.13
Past Tense	ed	8.03
Person/Object	er, or, men, person, ess, ist	7.23
Present Participle	ing	14.03
Slang	bros, bike, bbq, diff, h2o	0.12
Third Person	s	6.16
Time	fore, pre, post, re	0.54

Note.

Table 5

	Root	Raw	Affix	Previous COS	JCN	LSA	FSG	BSG
Root	1							
Raw	0.93	1						
Affix	0.5	0.53	1					
Previous COS	0.94	0.91	0.49	1				
JCN	-0.18	-0.22	-0.17	-0.22	1			
LSA	0.18	0.15	0.1	0.21	-0.06	1		
FSG	0.06	0.04	0.08	0.1	-0.15	0.24	1	
BSG	0.14	0.15	0.17	0.18	-0.18	0.26	0.31	1

Note.

Table 6

Percent and Mean Overlap between cue-feature lists from the free association norms

	% Overlap	*M* FSG	*SD* FSG	Min	Max
Adjective	51.86	.12	.15	.01	.94
Noun	36.48	.11	.14	.01	.91
Verb	32.15	.11	.13	.01	.94
Other	44.44	.13	.18	.01	.88
Total	37.47	.11	.14	.01	.94